(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 March 2004 (25.03.2004)

PCT

(10) International Publication Number
WO 2004/024067 A2

(51) International Patent Classification[7]: A61K

(21) International Application Number: PCT/US2003/028199

(22) International Filing Date: 10 September 2003 (10.09.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/409,307	10 September 2002 (10.09.2002)	US
60/419,089	18 October 2002 (18.10.2002)	US

(71) Applicant (*for all designated States except US*): **VICAL INCORPORATED** [US/US]; 10390 Pacific Center Court, San Diego, CA 92121-4340 (US).

(72) Inventor; and
(75) Inventor/Applicant (*for US only*): **HERMANSON, Gary, G.** [US/US]; 3159 Via de Caballo, Encinitas, CA 92024 (US).

(74) Agents: **STEFFE, Eric, K.** et al.; Sterne, Kessler, Goldstein & Fox, P.L.L.C., 1100 New York Avenue, Washington, DC 20005 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:
— *as to the identity of the inventor (Rule 4.17(i)) for all designations*
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)*
— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations*
— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations*

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2004/024067 A2

(54) Title: CODON-OPTIMIZED POLYNUCLEOTIDE-BASED VACCINES AGAINST BACILLUS ANTHRACIS INFECTION

(57) Abstract: The invention is related to polynucleotide-based anthrax vaccines. In particular, the invention is plasmids operably encoding Bacillus anthracis antigens, in which the naturally-occurring coding regions for the B. anthracis antigens have been modified for improved translation in human or other mammalian cells through codon optimization. In certain embodiments, the coding regions are also modified so as to remove potential N-linked glycosylation sites. B. anthracis antigens which are useful in the invention include, but are not limited to protective antigen (PA), lethal factor (LF), and fragments, variants or derivatives of either of these antigens. The invention is further directed to methods to induce an immune response to B. anthracis in a mammal, for example, a human, comprising delivering a plasmid encoding a codon-optimized B. anthracis antigen as described above. The invention is also directed to pharmaceutical compositions comprising plasmids encoding a codon-optimized B. anthracis antigen as described above, and further comprising adjuvants, excipients, or immune modulators.

CODON-OPTIMIZED POLYNUCLEOTIDE-BASED VACCINES AGAINST *BACILLUS ANTHRACIS* INFECTION

BACKGROUND OF THE INVENTION

Field of the Invention

Historically, anthrax infection is associated with herd animals and was not commonly seen as a human pathogen (Mock, M. and Fouet, A. *Annual Review of Microbiology* *55*:647-671(2001)). Therefore, it is not surprising that zoonotic *Bacillus anthracis* infection and pathogenesis in humans is not well characterized. However, anthrax has become a greater human disease problem with the realization that anthrax spores could be weaponized. It is now widely accepted that *B. anthracis* spores can be inexpensively produced, are extremely stable when properly stored, and could be effectively distributed in populated areas. Consequently, *B. anthracis* becomes an ideal organism for use as a biological weapon and opens up the possibility of an intentional and major outbreak of infection in humans. Research during the past 10-15 years has provided an increasing amount of information about the molecular basis of disease in humans, providing the scientific basis for developing specific diagnostics and defined subunit vaccines.

Related Art

In addition to developing more rapid and sensitive diagnostics, molecular biological methods enable the development of defined subunit vaccines to counter bioterrorism. Indeed, safe, effective recombinant subunit vaccines would significantly reduce, and perhaps eliminate, the need for therapeutic treatments. In the case of *B. anthracis*, virulence is the results of a multi-component toxin secreted by the organism. The toxin consists of three separate gene products designated protective antigen (PA), lethal factor (LF) and edema factor (EF). The genes encoding these toxin components (pag, lef, and cya, respectively) are located on a 184-kb plasmid designated pXO1.

pXO1, along with a second plasmid, pXO2 carrying capsule genes thought to protect bacilli from host cell phagocytosis, are required for full anthrax virulence and are carried by all virulent strains of *B. anthracis* (Mikesell, P., *et al. Infect. Immun. 39*: 371-376 (1983)). PA (735 aa, MW 82,684) is a single chain protein which binds to a mammalian cell surface receptor. Upon cleavage by furin (or a furin-like enzyme activity), it is cleaved into a 63-kDa receptor-bound product (Leppla, S.H., "Production and purification of anthrax toxin," in *Methods in Enzymology*. S. Harshman, ed., Academic Press, Inc., Orlando, FL (1988), pp. 103-116; Klimpel, K.R., *et al., Proc. Natl. Acad. Sci. (USA) 89*:10277-10281 (1992); Gordon, V.M., *et al., Infect. Immun. 63*:82-87 (1995); Petosa, C., *et al., Nature 385*:8833-8838 (1997)). The 63-kDa PA fragment forms a heptameric complex on the mammalian cell surface which is capable of interacting with the 90-kDa LF protein and the 89-kDa EF protein, which are subsequently internalized (Milne, J.C., *et al., J. Biol. Chem. 269*:20607-20612 (1994); Petosa, C., *et al., Nature 385*:8833-8838 (1997)). LF (776 aa, MW 90,237) is a zinc metalloprotease that cleaves several isoforms of MAP kinase kinase (Mek1, Mek2, MKK3), thereby disrupting signal transduction events within the cell and eventually leading to cell death (Duesbery, N.S., *et al., Science 280*:734-737 (1998); Pellizari, R., *et al., FEBS Ltrs 462*:199-204. (1999)). The EF protein (767 aa, MW 88,808) is a calmodulin-dependent adenylate cyclase that causes deregulation of cellular physiology, leading to clinical manifestations that include edema (Leppla, S.H., *Proc. Natl. Acad. Sci. (USA) 79*:3162-3166 (1982)). The LF protein, which together with PA is referred to as lethal toxin (Letx), is considered responsible for the rapid lethality of anthrax infection (Pannifer, A., *et al., Nature 414*:229-232. (2001)).

Protection against anthrax infection is associated with a humoral immune response directed against PA (Ivins, B.E. and Welkos, S.L., Eur. *J. Epidemiol. 4*(1):12-19 (1988); Ivins, B., *et al., Vaccine 13*:1779-1784 (1995)). Some evidence suggests that EF and LF may also contribute to specific immunity (Little, S.F. and Knudson, G.B., *Infect. Immun. 52*:509-512. (1986);

Ivins, B.E. and Welkos, S.L., *Eur. J. Epidemiol. 4*(1):12-19 (1988); Pezard, C., *et al., Infect. Immun. 63*:1369-1372 (1995)), although these components have not been formulated into a subunit vaccine.

The current FDA-approved anthrax vaccine, Anthrax Vaccine Adsorbed (AVA), is produced from the culture supernatant fraction of the V770-NP1-R strain of *B. anthracis*. Its principal component is the PA antigen adsorbed onto aluminum hydroxide. The production process is complex and the precise composition of the bacterial cell supernatant is not well characterized. Consequently, there is a significant lot-to-lot variation. In addition, the approved vaccination regimen is less than optimal for compliance and convenience: AVA is administered subcutaneously in a 0.5 ml volume, at 0, 2, and 4 weeks and then again at 6, 12, and 18 months. Annual boosts are also required.

Recently there has been a report of potential safety concerns in pregnant women, although the causal relationship has not been well established. As a result of these and other lay press reports, there is a negative public perception about the reliability and quality of the AVA vaccine even though the actual safety of the vaccine has never been seriously questioned in the scientific literature. A major concern with the current AVA anthrax vaccine is the paucity of analytical characterization of the actual composition of the vaccine preparation. It has been suggested that the presence of minute amounts of unspecified components <u>may</u> contribute to the adverse events that have been associated with administration of the AVA vaccine. In contrast, DNA vaccines are designed to elicit immunity against discrete, well-defined target antigens and are unlikely to be the subject of the same criticism. In short, DNA vaccines can be multivalent and yet highly defined.

During the past few years there has been substantial interest in testing DNA-based vaccines for a number of infectious diseases where the need for a vaccine, or an improved vaccine, exists. Several well-recognized advantages of DNA-based vaccines include the speed, ease and cost of manufacture, the versatility of developing and testing multivalent vaccines, the finding that

DNA vaccines can produce a robust cellular response in a wide variety of animal models as well as in man, and the proven safety of using plasmid DNA as a delivery vector (Donnelly, J.J., *et al., Annu. Rev. Immunol. 15*:617-648 (1997); Manickan, E., *et al., Crit. Rev. Immunol. 17*(2):139-154 (1997)). DNA vaccines represent the next generation in the development of vaccines (Nossal, G., *Nat. Med. 4*(5 Supple):475-476 (1998)) and numerous DNA vaccines are in clinical trials.

DNA-based immunization have already been shown, in animal models, to protect against a lethal challenge of anthrax toxin. The initial published work indicated that a plasmid encoding the protease-cleaved fragment (PA_{63}) of PA (Gu, *et al,. Vaccine 17*:340-344 (1999)) elicited protective immunity against a lethal toxin challenge. Price, *et al., Infection and Immunity 69*:4509-4515 (2001) extended these observations and demonstrated that DNA-based immunization with a fragment of the LF gene product would also contribute to or provide protection against a lethal toxin challenge. Having established proof of principle in pre-clinical studies, we now propose to develop an aggressive product development plan that will lead to an efficacious human vaccine against *B. anthracis* using a DNA-based immunization strategy.

Retooling coding regions encoding polypeptides from pathogens using codon frequencies preferred in a given mammalian species often results in a significant increase in expression in the cells of that mammalian species, and concomitant increase in immunogenicity. *See, e.g.*, Deml, L., *et al., J. Virol. 75*:10991-11001 (2001), and Narum, DL, *et al., Infect. Immun. 69*:7250-7253 (2001).

There remains a need in the art for convenient, safe, and efficacious immunogenic compounds to protect vertebrates against *Bacillus anthracis* infection. The present invention provides safe yet effective immunogenic compounds and methods to protect vertebrates against *Bacillus anthracis* infection using such immunogenic compounds.

SUMMARY OF THE INVENTION

The present invention is directed to enhancing immune response of a vertebrate in need of protection against anthrax infection by administering *in vivo*, into a tissue of a vertebrate, a polynucleotide comprising a codon-optimized coding region encoding a component of *Bacillus anthracis* lethal toxin or nucleic acid fragments of such coding regions encoding fragments, variants, or derivatives thereof. Nucleic acid fragments of the present invention are altered from their native state in one or more of the following ways. First, a nucleic acid fragment which encodes a component of the *B. anthracis* lethal toxin may be part or all of a codon-optimized coding region, optimized according to codon usage in a given species, *e.g.*, a vertebrate species, *e.g.*, a mammalian species, *e.g.*, humans. In addition, a nucleic acid fragment which encodes a component of the *B. anthracis* lethal toxin may be a fragment which encodes only a portion of a full-length polypeptide, and/or may be mutated so as to, for example, remove from the encoded polypeptide adventitious protein motifs present in the encoded polypeptide or virulence factors associated with the encoded polypeptide. For example, the nucleic acid sequence could be mutated so as not to encode adventitious N-linked glycosylation motifs (N-X-(S or T), where X is any amino acid). The polynucleotides are incorporated into the cells of the vertebrate *in vivo*, and a prophylactically or therapeutically effective amount of a *Bacillus anthracis* lethal toxin component is produced *in vivo*.

The invention further provides immunogenic compositions comprising a polynucleotide which comprises one or more codon-optimized coding regions encoding components of *Bacillus anthracis* lethal toxin or nucleic acid fragments of such coding regions encoding fragments, variants, or derivatives thereof, and methods for enhancing the immune response of a vertebrate to *Bacillus anthracis* infection by administering to the tissues of a vertebrate one or more polynucleotides comprising one or more codon-optimized coding regions encoding components of *Bacillus anthracis* lethal toxin or nucleic acid

fragments of such coding regions encoding fragments, variants, or derivatives thereof. The present invention further provides plasmids and other polynucleotide constructs for delivery of nucleic acid coding sequences to a vertebrate which provide expression of *Bacillus anthracis* toxin components, or fragments, variants, or derivatives thereof.

In certain embodiments, the invention further provides methods for enhancing the immune response of a vertebrate to *Bacillus anthracis* infection by sequentially administering two or more different immunogenic compositions to the tissues of the vertebrate. Such methods comprise initially administering one or more polynucleotides comprising one or more codon-optimized coding regions encoding components of *Bacillus anthracis* lethal toxin or nucleic acid fragments of such coding regions encoding fragments, variants, or derivatives thereof, to prime immunity, and then administering subsequently a different vaccine composition, for example a recombinant viral vaccine, a protein subunit vaccine, or a recombinant or killed bacterial vaccine or vaccines to boost the anti-*Bacillus anthracis* toxin immune response in the vertebrate.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the nucleotide sequence (SEQ ID NO:1) and amino acid translation (SEQ ID NO:2) of TPA-PA63. SEQ ID NO:1 contains a nucleic acid fragment of a human codon-optimized PA coding region, encoding the 63kD furin cleavage product of the *Bacillus anthracis* protective antigen (PA), fused to a nucleic acid encoding the human tissue plasminogen activator (TPA) signal peptide sequence. Nucleotides 1-12 of SEQ ID NO:1 is a Kozak translation initiation element and nucleotides 13-81 of SEQ ID NO:1 encode the TPA signal peptide. Nucleotides 82-1782 of SEQ ID NO:1 encode the 63kD furin processed fragment of PA that can bind LF and EF, and heptamerize and form a pore in infected cells through which the toxin is delivered. The 63kD furin processed fragment of PA corresponds to amino

acids 199-764 of the native full-length PA amino acid sequence of GenBank accession No. AAA2263 (SEQ ID NO:4) encoded by GenBank accession number M22589 (SEQ ID NO:3).

Figure 2 shows the nucleotide sequence (SEQ ID NO:5) and amino acid translation (SEQ ID NO:6) of TPA-PA63ΔF313-314. SEQ ID NO:5 is identical to SEQ ID NO:1, except that the nucleotides encoding the two phenylalanine residues at amino acids 313-314 of SEQ ID NO:2 are deleted, which results in a PA protein that cannot form the pore through which LF and EF are translocated. Nucleotides 1-12 of SEQ ID NO:5 is a Kozak translation initiation element and nucleotides 13-81 of SEQ ID NO:5 encode the TPA signal peptide.

Figure 3 shows the nucleotide sequence (SEQ ID NO:7) and amino acid translation (SEQ ID NO:8) encoding TPA-PA83 Δ Furin. SEQ ID NO:7 contains a nucleic acid fragment of a human codon-optimized PA coding region, encoding full-length mature PA (amino acids 30-764 of SEQ ID NO:4) with the furin cleavage site deleted (SRKKRS, amino acids 192-197 of SEQ ID NO:4). This mutant PA cannot be processed to the 63 kD fragment and cannot bind LF or EF. Nucleotides 1-12 of SEQ ID NO:7 is a Kozak translation initiation element and nucleotides 13-81 of SEQ ID NO:7 encode the TPA signal peptide.

Figure 4 shows the nucleotide sequence (SEQ ID NO:9) and amino acid translation (SEQ ID NO:10) of TPA-LF HEXXH (H686A+H690A+E687D). SEQ ID NO:9 contains a nucleic acid fragment of a human codon-optimized LF coding region, encoding the mature *Bacillus anthracis* lethal factor with three inactivating point mutations. Either the H686A + H690A (decreased Zn binding and no protease activity) or E687D (no protease activity, no *in vitro* or *in vivo* macrophage killing) mutation inactivates the enzymatic activity of LF rendering it non-toxic (Hammond S.E. and Hanna P.C. *Infect Immun.* *66*:2374-2378(1998)). This construct combines both sets of mutations. Nucleotides 1-12 of SEQ ID NO:9 is a Kozak translation initiation element and nucleotides 13-81 of SEQ ID NO:9

encode the TPA signal peptide. Nucleotides 82- 2412 encode a non-toxic form of lethal factor. TPA-LF HEXXH (H686A+H690A+E687D) is derived from the native full-length LF amino acid sequence of GenBank accession No. AAA22569 (SEQ ID NO:12) encoded by GenBank accession number M30210 (SEQ ID NO:11).

Figure 5 shows the nucleotide sequence (SEQ ID NO:13) and amino acid translation (SEQ ID NO:14) of TPA-LF Domain I-III. SEQ ID NO:13 contains a nucleic acid fragment of a human codon-optimized LF coding region, encoding an N-terminal fragment (domains I-III) of LF corresponding to amino acids 34-583 of SEQ ID NO:12. Nucleotides 1-12 of SEQ ID NO:13 is a Kozak translation initiation element and nucleotides 13-81 of SEQ ID NO:13 encode the TPA signal peptide. Nucleotides 82-1734 of SEQ ID NO:13 encode domains I-III of LF. The entire protease domain (domain IV) has been deleted.

Figure 6 shows the nucleotide sequence (SEQ ID NO:15) and amino acid translation (SEQ ID NO:16) of TPA-LF Domain IA. SEQ ID NO:15 contains a nucleic acid fragment of a human codon-optimized LF coding region, encoding an LF N-terminal fragment of LF corresponding to amino acids 34-254 of SEQ ID NO:12. This truncated version of LF roughly corresponds to the domain I portion of LF that directly binds PA63. Pannifer A.D. *et al.* Nature 414:229-333 (2001). Nucleotides 1-12 of SEQ ID NO:15 is a Kozak translation initiation element and nucleotides 13-81 of SEQ ID NO:15 encode the TPA signal peptide. Nucleotides 82-747 of SEQ ID NO:15 encode domain I of LF.

Figure 7 shows the nucleotide sequence (SEQ ID NO:17) and amino acid translation (SEQ ID NO:18) of TPA-PA63 with the N-linked glycosylation motifs mutated. SEQ ID NO:17 is identical to SEQ ID NO:1, except that all ten N-linked glycosylation sites have been mutated. The N residue in the glycosylation motif (N-X-S/T) has been changed to a Q residue (Q-X-S/T) resulting in a protein that cannot glycosylated at these sites. Nucleotides 1-12 of SEQ ID NO:17 is a Kozak translation initiation element

and nucleotides 13-81 of SEQ ID NO:17 encode the TPA signal peptide. Nucleotides 82-747 of SEQ ID NO:15 encode domain I of LF. Nucleotides 82-1782 of SEQ ID NO:17 encode a mutated form of the 63kD furin processed fragment of PA that can heptamerize, bind LF and EF, and form a pore in infected cells through which the toxin is delivered.

Figure 8 shows the nucleotide sequence (SEQ ID NO:19) and amino acid translation (SEQ ID NO:20) of sugar-minus TPA-LF HEXXH mutant (H686A+H690A+E687D). SEQ ID NO:19 is identical to SEQ ID NO:9, except that all seven N-linked glycosylation sites have been mutated. The N residue in the glycosylation motif (N-X-S/T) has been changed to a Q residue (Q-X-S/T) resulting in a protein that cannot be glycosylated at these sites. Nucleotides 1-12 of SEQ ID NO:19 is a Kozak translation initiation element and nucleotides 13-81 of SEQ ID NO:19 encode the TPA signal peptide. Nucleotides 82- 2412 encode a non-toxic form of lethal factor which cannot be glycosylated.

Figure 9 shows a nucleotide sequence comparison of a nucleic acid fragment of a human codon-optimized PA coding region, encoding PA63 (nucleotides 82-1782 of SEQ ID NO:1) vs. the native nucleotide sequence of *Bacillus anthracis* PA63 (nucleotides 2398-4095 of SEQ ID NO:3). Differences between the two sequences are denoted with a letter. There is approximately 25% difference in the two coding sequences.

Figure 10 shows a nucleotide sequence comparison of a humanized nucleotide sequence encoding the mature PA Δ furin (nucleotides 82-2268 of SEQ ID NO: 7) vs. the native nucleotide sequence of *Bacillus anthracis* mature PA (nucleotides 1891-4095 of SEQ ID NO:3). Differences between the two sequences are denoted with a letter and gaps are denoted as a dash. There is approximately 25% difference in the two coding sequences.

Figure 11 shows a nucleotide sequence comparison of a humanized nucleotide sequence encoding the mature LF Δ HEXXH (nucleotides 82-2409 of SEQ ID NO:9) vs. the native nucleotide sequence of *Bacillus anthracis* mature LF (nucleotides 784-3111 of SEQ ID NO:11). Differences between

the two sequences are denoted with a letter and gaps are denoted by a gap. There is approximately 25% difference in the two coding sequences.

Figure 12 shows an amino acid comparison between TPA-PA63 (SEQ ID NO:2) and sugar minus TPA-PA63 (SEQ ID NO:18). All ten N-linked glycosylation sites N-X-S/T in TPA-PA63 have been mutated to Q-X-S/T so that they will not be a substrate for glycosylation.

Figure 13 shows an amino acid comparison between TPA-LFΔHEXXH (SEQ ID NO:10) and sugar minus TPA-LFΔHEXXH (SEQ ID NO:20). All seven N-linked glycosylation sites N-X-S/T in TPA-PA63 have been mutated to Q-X-S/T so that they will not be a substrate for glycosylation.

Figure 14 shows the nucleotide sequence (SEQ ID NO:39) and amino acid translation (SEQ ID NO:40) of TPA-LF Domain IB. SEQ ID NO:39 contains a nucleic acid fragment of a human codon-optimized LF coding region, encoding an LF N-terminal fragment of LF corresponding to amino acids 34-295 of SEQ ID NO:12. This truncated version of LF roughly corresponds to the domain I portion of LF that directly binds PA63. Pannifer A.D. *et al.* Nature 414:229-333 (2001). Nucleotides 1-12 of SEQ ID NO:39 is a Kozak translation initiation element and nucleotides 13-81 of SEQ ID NO:39 encode the TPA signal peptide. Nucleotides 82-870 of SEQ ID NO:39 encode domain I of LF.

Figure 15: Antibody titers measured in mouse immunization experiment 1 (Example 11). 15A: protective antigen (PA) titers; 15B: lethal factor (LF) titers; and 15C: lethal toxin (LT) neutralization titers.

Figure 16: Antibody titers measured in mouse immunization experiment 2 (Example 11). 16A: protective antigen (PA) titers; 16B: lethal factor (LF) titers; and 16C: lethal toxin (LT) neutralization titers.

Figure 17: Antibody titers measured in mouse immunization experiment 3 (Example 11). 17A: protective antigen (PA) titers; 17B: lethal factor (LF) titers; and 17C: lethal toxin (LT) neutralization titers.

Figure 18: Antibody titers measured in mouse immunization experiment 4 (Example 11). 18A: protective antigen (PA) titers; 185B: lethal toxin (LT) neutralization titers.

Figure 19: Pre-challenge lethal toxin (LT) neutralization titers in the rabbit immunization experiment (Example 12).

Figure 20: Antibody titers measured in mouse immunization experiment 5 (Example 11).

Figure 21: Lethal toxin (LT) neutralization titers in mouse immunization experiment 5 (Example 11).

DETAILED DESCRIPTION OF THE INVENTION

The present invention is directed to compositions and methods for enhancing the immune response of a vertebrate in need of protection against *Bacillus anthracis* infection by administering *in vivo*, into a tissue of a vertebrate, a polynucleotide comprising a human codon-optimized coding region encoding a polypeptide of *Bacillus anthracis*, or a nucleic acid fragment of such a coding region encoding a fragment, variant, or derivative thereof. The polynucleotides are incorporated into the cells of the vertebrate *in vivo*, and an immunologically effective amount of the *Bacillus anthracis* polypeptide, or fragment or variant is produced *in vivo*.

The present invention provides polynucleotide-based vaccines and methods for delivery of *Bacillus anthracis* coding sequences to a vertebrate with optimal expression and safety conferred through codon optimization and/or other manipulations. These polynucleotide-based vaccines are prepared and administered in such a manner that the encoded gene products are optimally expressed in the particular vertebrate to which the composition is administered. As a result, these compositions and methods are useful in stimulating an immune response against *Bacillus anthracis* infection as the coding sequence encodes a polypeptide which stimulates the immune system to respond to anthrax infection. Also included in the invention are expression

systems, delivery systems, and codon-optimized *Bacillus anthracis* coding sequences.

A polynucleotide vaccine of the present invention is capable of eliciting, without more, an immune response in a vertebrate against *B. anthracis* when administered to that vertebrate. Such polynucleotides are referred to herein as polynucleotide vaccines.

It is to be noted that the term "a" or "an" entity refers to one or more of that entity; for example, "a polynucleotide," is understood to represent one or more polynucleotides. As such, the terms "a" (or "an"), "one or more," and "at least one" can be used interchangeably herein.

The terms "nucleic acid" or "nucleic acid fragment" refers to any one or more nucleic acid segments, *e.g.*, DNA or RNA fragments, present in a polynucleotide or construct. While the terms "nucleic acid," as used herein, is meant to include any nucleic acid, the term "nucleic acid fragment" is used herein to specifically denote a fragment of a designed or synthetic codon-optimized coding region encoding a polypeptide, or fragment, variant, or derivative thereof, which has been optimized according to the codon usage of a given species. As used herein, a "coding region" is a portion of nucleic acid which consists of codons translated into amino acids. Although a "stop codon" (TAG, TGA, or TAA) is not translated into an amino acid, it may be considered to be part of a coding region, but any flanking sequences, for example promoters, ribosome binding sites, transcriptional terminators, and the like, are not part of a coding region. Two or more nucleic acids of the present invention can be present in a single polynucleotide construct, *e.g.*, on a single plasmid, or in separate polynucleotide constructs, *e.g.*, on separate plasmids. Furthermore, any nucleic acid or nucleic acid fragment may encode a single polypeptide, *e.g.*, a single antigen, cytokine, or regulatory polypeptide, or may encode more than one polypeptide, *e.g.*, a nucleic acid may encode two or more polypeptides. In addition, a nucleic acid may encode a regulatory element such as a promoter or a transcription terminator, or may

encode a specialized element or motif of a polypeptide or protein, such as a secretory signal peptide or a functional domain.

The terms "fragment," "variant," "derivative" and "analog" when referring to *B. anthracis* polypeptides of the present invention include any polypeptides which retain at least some of the immunogenicity or antigenicity of the corresponding native polypeptide. Fragments of *B. anthracis* polypeptides of the present invention include proteolytic fragments, deletion fragments and in particular, fragments of *B. anthracis* polypeptides which exhibit reduced pathogenicity when delivered to an animal. Polypeptide fragments further include any portion of the polypeptide which comprises an antigenic or immunogenic epitope of the native polypeptide, including linear as well as three-demensional epitopes. Variants of *B. anthracis* polypeptides of the present invention includes fragments as described above, and also polypeptides with altered amino acid sequences due to amino acid substitutions, deletions, or insertions. Variants may occur naturally, such as an allelic variant. By an "allelic variant" is intended alternate forms of a gene occupying a given locus on a chromosome of an organism. *Genes II*, Lewin, B., ed., John Wiley & Sons, New York (1985). Non-naturally occurring variants may be produced using art-known mutagenesis techniques. Variant polypeptides may comprise conservative or non-conservative amino acid substitutions, deletions or additions. Derivatives of *B. anthracis* polypeptides of the present invention, are polypeptides which have been altered so as to exhibit additional features not found on the native polypeptide. Examples include fusion proteins. An analog is another form of a *B. anthracis* polypeptide of the present invention, An example is a proprotein (*e.g.*, *B. anthracis* PA83) which can be activated by cleavage of the proprotein to produce an active mature polypeptide (*e.g.*, *B. anthracis* PA63).

The term "polynucleotide" is intended to encompass a singular nucleic acid or nucleic acid fragment as well as plural nucleic acids or nucleic acid fragments, and refers to an isolated molecule or construct, *e.g.*, a virus genome (*e.g.*, a non-infectious viral genome), messenger RNA (mRNA), plasmid DNA

(pDNA), or derivatives of pDNA (*e.g.*, minicircles as described in (Darquet, A-M *et al.*, *Gene Therapy 4*:1341-1349 (1997)) comprising a polynucleotide. A nucleic acid may be provided in linear (*e.g.*, mRNA), circular (*e.g.*, plasmid), or branched form as well as double-stranded or single-stranded forms. A polynucleotide may comprise a conventional phosphodiester bond or a non-conventional bond (*e.g.*, an amide bond, such as found in peptide nucleic acids (PNA)).

The terms "infectious polynucleotide" or "infectious nucleic acid" are intended to encompass isolated viral polynucleotides and/or nucleic acids which are solely sufficient to mediate the synthesis of complete infectious virus particles upon uptake by permissive cells. "Isolated" means that the viral nucleic acid does not require pre-synthesized copies of any of the polypeptides it encodes, *e.g.*, viral replicases, in order to initiate its replication cycle.

The terms "non-infectious polynucleotide" or "non-infectious nucleic acid" as defined herein which cannot, without additional added materials, e.g, polypeptides, mediate the synthesis of complete infectious virus particles upon uptake by permissive cells. An infectious polynucleotide or nucleic acid is not made "non-infectious" simply because it is taken up by a non-permissive cell. For example, an infectious viral polynucleotide from a virus with limited host range is infectious if it is capable of mediating the synthesis of complete infectious virus particles when taken up by cells derived from a permissive host (*i.e.*, a host permissive for the virus itself). The fact that uptake by cells derived from a non-permissive host does not result in the synthesis of complete infectious virus particles does not make the nucleic acid "non-infectious." In other words, the term is not qualified by the nature of the host cell, the tissue type, or the species.

In some cases, an isolated infectious polynucleotide or nucleic acid may produce fully-infectious virus particles in a host cell population which lacks receptors for the virus particles, *i.e.*, is non-permissive for the virus itself. Thus viruses produced will not infect surrounding cells. However, if

the supernatant containing the virus particles is transferred to cells which are permissive for the virus, infection will take place.

The terms "replicating polynucleotide" or "replicating nucleic acid" are meant to encompass those polynucleotides and/or nucleic acids which, upon being taken up by a permissive host cell, are capable of producing multiple, *e.g.*, one or more copies of the same polynucleotide or nucleic acid. Infectious polynucleotides and nucleic acids are a <u>subset</u> of replicating polynucleotides and nucleic acids; the terms are not synonymous. For example, a defective virus genome lacking the genes for virus coat proteins may replicate, *e.g.*, produce multiple copies of itself, but is NOT infectious because it is incapable of mediating the synthesis of complete infectious virus particles unless the coat proteins, or another nucleic acid encoding the coat proteins, are provided.

In certain embodiments, the polynucleotide, nucleic acid, or nucleic acid fragment is DNA. In the case of DNA, a polynucleotide comprising a nucleic acid which encodes a polypeptide normally also comprises a promoter operably associated with the polypeptide-encoding nucleic acid. An operable association is when a nucleic acid encoding a gene product, *e.g.*, a polypeptide, is associated with one or more regulatory sequences in such a way as to place expression of the gene product under the influence or control of the regulatory sequence(s). Two DNA fragments (such as a polypeptide-encoding nucleic acid and a promoter associated with the 5' end of the nucleic acid) are "operably associated" if induction of promoter function results in the transcription of mRNA encoding the desired gene product and if the nature of the linkage between the two DNA fragments does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the expression regulatory sequences to direct the expression of the gene product, or (3) interfere with the ability of the DNA template to be transcribed. Thus, a promoter region would be operably associated with a nucleic acid encoding a polypeptide if the promoter was capable of effecting transcription of that nucleic acid. The promoter may be a cell-specific promoter that directs substantial transcription of the DNA only in predetermined cells. Other

transcription control elements, besides a promoter, for example enhancers, operators, repressors, and transcription termination signals, can be operably associated with the polynucleotide to direct cell-specific transcription. Suitable promoters and other transcription control regions are disclosed herein.

A variety of transcription control regions are known to those skilled in the art. These include, without limitation, transcription control regions which function in vertebrate cells, such as, but not limited to, promoter and enhancer segments from cytomegaloviruses (the immediate early promoter, in conjunction with intron-A), simian virus 40 (the early promoter), retroviruses (such as Rous sarcoma virus), and picornaviruses (particularly an internal ribosome entry site, or IRES, also referred to as a CITE sequence). Other transcription control regions include those derived from vertebrate genes such as actin, heat shock protein, bovine growth hormone and rabbit β-globin, as well as other sequences capable of controlling gene expression in eukaryotic cells. Additional suitable transcription control regions include tissue-specific promoters and enhancers as well as lymphokine-inducible promoters (*e.g.*, promoters inducible by interferons or interleukins).

In one embodiment, a DNA polynucleotide of the present invention is a circular or linearized plasmid, or other linear DNA which is, in certain embodiments, non-infectious and nonintegrating (*i.e.*, does not integrate into the genome of vertebrate cells). A linearized plasmid is a plasmid that was previously circular but has been linearized, for example, by digestion with a restriction endonuclease.

Alternatively, DNA virus genomes may be used to administer DNA polynucleotides into vertebrate cells. In certain embodiments, a DNA virus genome of the present invention is noninfectious, and nonintegrating. Suitable DNA virus genomes include herpesvirus genomes, adenovirus genomes, adeno-associated virus genomes, and poxvirus genomes. References citing methods for the *in vivo* introduction of non-infectious virus genomes to vertebrate tissues are well known to those of ordinary skill in the art, and are cited *supra*.

In other embodiments, a polynucleotide of the present invention is RNA. In a suitable embodiment, the RNA is in the form of messenger RNA (mRNA). Methods for introducing RNA sequences into vertebrate cells are described in U.S. Patent No. 5,580,859, the disclosure of which is incorporated herein by reference in its entirety.

Polynucleotide, nucleic acids, and nucleic acid fragments of the present invention may be associated with additional nucleic acids which encode secretory or signal peptides, which direct the secretion of a polypeptide encoded by a nucleic acid or polynucleotide of the present invention. According to the signal hypothesis, proteins secreted by mammalian cells have a signal peptide or secretory leader sequence which is cleaved from the mature protein once export of the growing protein chain across the rough endoplasmic reticulum has been initiated. Those of ordinary skill in the art are aware that polypeptides secreted by vertebrate cells generally have a signal peptide fused to the N-terminus of the polypeptide, which is cleaved from the complete or "full length" polypeptide to produce a secreted or "mature" form of the polypeptide. In certain embodiments, the native leader sequence is used, or a functional derivative of that sequence that retains the ability to direct the secretion of the polypeptide that is operably associated with it. Alternatively, a heterologous mammalian leader sequence, or a functional derivative thereof, may be used. For example, the wild-type leader sequence may be substituted with the leader sequence of human tissue plasminogen activator (TPA) or mouse β-glucuronidase.

In accordance with one aspect of the present invention, there is provided a plasmid for expression of a *Bacillus anthracis* PA or LF-derived coding sequence optimized for expression in the particular vertebrate species to be treated or immunized. When such a plasmid is delivered, *in vivo* to a tissue of the vertebrate to be treated or immunized, the transcriptional unit will thus express the encoded gene product. The level of expression of the gene product will depend to a significant extent on the strength of the associated

promoter and the presence and activation of an associated enhancer element, as well as the optimization of the coding region.

As used herein, the term "plasmid" refers to a construct made up of genetic material (*i.e.*, nucleic acids). Typically a plasmid contains an origin of replication which is functional in bacterial host cells, *e.g.*, *Eschericha coli*, and selectable markers for detecting bacterial host cells comprising the plasmid. Plasmids of the present invention may include genetic elements as described herein arranged such that an inserted coding sequence can be transcribed in eukaryotic cells. Also, while the plasmid may include a sequence from a viral nucleic acid, such viral sequence normally does not cause the incorporation of the plasmid into a viral particle, and the plasmid is therefore a non-viral vector. In certain embodiments described herein, a plasmid is a closed circular DNA molecule.

The term "expression" refers to the biological production of a product encoded by a coding sequence. In most cases a DNA sequence, including the coding sequence, is transcribed to form a messenger-RNA (mRNA). The messenger-RNA is translated to form a polypeptide product which has a relevant biological activity. Also, the process of expression may involve further processing steps to the RNA product of transcription, such as splicing to remove introns, and/or post-translational processing of a polypeptide product.

As used herein, the term “polypeptide” is intended to encompass a singular “polypeptide” as well as plural “polypeptides,” and comprises any chain or chains of two or more amino acids. Thus, as used herein, terms including, but not limited to “peptide,” “dipeptide,” “tripeptide,” “protein,” “amino acid chain,” or any other term used to refer to a chain or chains of two or more amino acids, are included in the definition of a “polypeptide,” and the term “polypeptide” may be used instead of, or interchangeably with any of these terms. The term further includes polypeptides which have undergone post-translational modifications, for example, glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking

groups, proteolytic cleavage, or modification by non-naturally occurring amino acids.

Also included as polypeptides of the present invention are fragments, derivatives, analogs, or variants of the foregoing polypeptides, and any combination thereof. Polypeptides, and fragments, derivatives, analogs, or variants thereof of the present invention can be antigenic and immunogenic polypeptides related to *B. anthracis* polypeptides, which are used to prevent or treat, *i.e.*, cure, ameliorate, lessen the severity of, or prevent or reduce contagion of infectious disease caused by *B. anthracis*.

As used herein, an antigenic polypeptide or an immunogenic polypeptide is a polypeptide which, when introduced into a vertebrate, reacts with the immune system molecules of the vertebrate, *i.e.*, is antigenic, and/or induces an immune response in the vertebrate, *i.e.*, is immunogenic. It is quite likely that an immunogenic polypeptide will also be antigenic, but an antigenic polypeptide, because of its size or conformation, may not necessarily be immunogenic. Examples of antigenic and immunogenic polypeptides of the present invention include, but are not limited to, *B. anthracis* protective antigen (PA) or lethal factor (LF), fragments thereof, *e.g.*, PA63, LF domains I-III or domain I, variants thereof, *e.g.*, PA63Δ FF, PA83 Δ furin, PA63 sugar minus, LF HEXXH, or LF sugar minus (all described in more detail herein) and derivatives thereof, *e.g.*, any of the foregoing polypeptides fused to a TPA signal peptide.

The term "epitopes," as used herein, refers to portions of a polypeptide having antigenic or immunogenic activity in an animal, for example a mammal, for example, a human. An "immunogenic epitope," as used herein, is defined as a portion of a protein that elicits an immune response in an animal, as determined by any method known in the art. The term "antigenic epitope," as used herein, is defined as a portion of a protein to which an antibody can immunospecifically bind its antigen as determined by any method well known in the art. Immunospecific binding excludes non-specific binding but does not

necessarily exclude cross-reactivity with other antigens. Antigenic epitopes need not necessarily be immunogenic.

In the present invention, antigenic epitopes preferably contain a sequence of at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, or between about 15 to about 30 amino acids contained within the amino acid sequence of a polypeptide of the invention. Certain polypeptides comprising immunogenic or antigenic epitopes are at least 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 amino acid residues in length. Antigenic as well as immunogenic epitopes may be linear, *i.e.*, be comprised of contiguous amino acids in a polypeptide, or may be three dimensional, *i.e.*, where an epitope is comprised of non-contiguous amino acids which come together due to the secondary or tertiary structure of the polypeptide, thereby forming an epitope.

The present invention is directed towards polynucleotides comprising nucleic acid fragments of codon-optimized coding regions which encode polypeptides of *Bacillus anthracis*, and in particular, *Bacillus anthracis* protective antigen (PA) or lethal factor (LF), and fragments, variants, or derivatives thereof.

"Codon optimization" is defined as modifying a nucleic acid sequence for enhanced expression in the cells of the vertebrate of interest by replacing at least one, more than one, or a significant number, of codons of the native sequence with codons that are more frequently or most frequently used in the genes of that vertebrate. Various species exhibit particular bias for certain codons of a particular amino acid.

The present invention relates to polynucleotides comprising nucleic acid fragments of codon-optimized coding regions which encode *Bacillus anthracis* polypeptides, with the codon usage adapted for optimized expression in the cells of a given vertebrate. These polynucleotides are prepared by incorporating codons preferred for use in the genes of a given species into the DNA sequence. Also provided are polynucleotide expression constructs, vectors, host cells comprising nucleic acid fragments of codon-

optimized coding regions which encode *Bacillus anthracis* polypeptides, and various methods of using the polynucleotide expression constructs, vectors, host cells to treat or prevent anthrax in a vertebrate.

Codon Optimization

As used herein the term "codon optimized coding region" means a nucleic acid coding region that has been adapted for expression in the cells of a given vertebrate by replacing at least one, or more than one, or a significant number, of codons with one or more codons that are more frequently used in the genes of that vertebrate.

Deviations in the nucleotide sequence that comprise the codons encoding the amino acids of any polypeptide chain allow for variations in the sequence coding for the gene. Since each codon consists of three nucleotides, and the nucleotides comprising DNA are restricted to four specific bases, there are 64 possible combinations of nucleotides, 61 of which encode amino acids (the remaining three codons encode signals ending translation). The "genetic code" which shows which codons encode which amino acids is reproduced herein as Table 1. As a result, many amino acids are designated by more than one codon. For example, the amino acids alanine and proline are coded for by four triplets, serine and arginine by six, whereas tryptophan and methionine are coded by just one triplet. This degeneracy allows for DNA base composition to vary over a wide range without altering the amino acid sequence of the proteins encoded by the DNA.

TABLE 1: The Standard Genetic Code

	T	C	A	G
T	TTT Phe (F) TTC " TTA Leu (L) TTG "	TCT Ser (S) TCC " TCA " TCG "	TAT Tyr (Y) TAC " TAA **Ter** TAG **Ter**	TGT Cys (C) TGC TGA **Ter** TGG Trp (W)
C	CTT Leu (L) CTC " CTA " CTG "	CCT Pro (P) CCC " CCA " CCG "	CAT His (H) CAC " CAA Gln (Q) CAG "	CGT Arg (R) CGC " CGA " CGG "
A	ATT Ile (I) ATC " ATA " **ATG** Met (M)	ACT Thr (T) ACC " ACA " ACG "	AAT Asn (N) AAC " AAA Lys (K) AAG "	AGT Ser (S) AGC " AGA Arg (R) AGG "
G	GTT Val (V) GTC " GTA " GTG "	GCT Ala (A) GCC " GCA " GCG "	GAT Asp (D) GAC " GAA Glu (E) GAG "	GGT Gly (G) GGC " GGA " GGG "

Many organisms display a bias for use of particular codons to code for insertion of a particular amino acid in a growing peptide chain. Codon preference or codon bias, differences in codon usage between organisms, is afforded by degeneracy of the genetic code, and is well documented among many organisms. Codon bias often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, *inter alia*, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization.

Given the large number of gene sequences available for a wide variety of animal, plant and microbial species, it is possible to calculate the relative

frequencies of codon usage. Codon usage tables are readily available, for example, at the "Codon Usage Database" available at http://www.kazusa.or.jp/codon/ (visited July 9, 2002), and these tables can be adapted in a number of ways. *See* Nakamura, Y., *et al.* "Codon usage tabulated from the international DNA sequence databases: status for the year 2000" *Nucl. Acids Res.* 28:292 (2000). Codon usage tables for human, mouse, domestic cat, and cow, calculated from GenBank Release 128.0 [15 February 2002], are reproduced below as Tables 2-5. These tables use mRNA nomenclature, and so instead of thymine (T) which is found in DNA, the tables use uracil (U) which is found in RNA. The tables have been adapted so that frequencies are calculated for each amino acid, rather than for all 64 codons.

TABLE 2: Codon Usage Table for Human Genes (*Homo sapiens*)

Amino Acid	Codon	Number	Frequency
Phe	UUU	326146	0.4525
Phe	UUC	394680	0.5475
Total		720826	
Leu	UUA	139249	0.0728
Leu	UUG	242151	0.1266
Leu	CUU	246206	0.1287
Leu	CUC	374262	0.1956
Leu	CUA	133980	0.0700
Leu	CUG	777077	0.4062
Total		1912925	
Ile	AUU	303721	0.3554
Ile	AUC	414483	0.4850
Ile	AUA	136399	0.1596
Total		854603	
Met	AUG	430946	1.0000
Total		430946	
Val	GUU	210423	0.1773
Val	GUC	282445	0.2380
Val	GUA	134991	0.1137

Amino Acid	Codon	Number	Frequency
Val	GUG	559044	0.4710
Total		1186903	
Ser	UCU	282407	0.1840
Ser	UCC	336349	0.2191
Ser	UCA	225963	0.1472
Ser	UCG	86761	0.0565
Ser	AGU	230047	0.1499
Ser	AGC	373362	0.2433
Total		1534889	
Pro	CCU	333705	0.2834
Pro	CCC	386462	0.3281
Pro	CCA	322220	0.2736
Pro	CCG	135317	0.1149
Total		1177704	
Thr	ACU	247913	0.2419
Thr	ACC	371420	0.3624
Thr	ACA	285655	0.2787
Thr	ACG	120022	0.1171
Total		1025010	
Ala	GCU	360146	0.2637
Ala	GCC	551452	0.4037
Ala	GCA	308034	0.2255
Ala	GCG	146233	0.1071
Total		1365865	
Tyr	UAU	232240	0.4347
Tyr	UAC	301978	0.5653
Total		534218	
His	CAU	201389	0.4113
His	CAC	288200	0.5887
Total		489589	
Gln	CAA	227742	0.2541
Gln	CAG	668391	0.7459
Total		896133	

Amino Acid	Codon	Number	Frequency
Asn	AAU	322271	0.4614
Asn	AAC	376210	0.5386
Total		698481	
Lys	AAA	462660	0.4212
Lys	AAG	635755	0.5788
Total		1098415	
Asp	GAU	430744	0.4613
Asp	GAC	502940	0.5387
Total		933684	
Glu	GAA	561277	0.4161
Glu	GAG	787712	0.5839
Total		1348989	
Cys	UGU	190962	0.4468
Cys	UGC	236400	0.5532
Total		427362	
Trp	UGG	248083	1.0000
Total		248083	
Arg	CGU	90899	0.0830
Arg	CGC	210931	0.1927
Arg	CGA	122555	0.1120
Arg	CGG	228970	0.2092
Arg	AGA	221221	0.2021
Arg	AGG	220119	0.2011
Total		1094695	
Gly	GGU	209450	0.1632
Gly	GGC	441320	0.3438
Gly	GGA	315726	0.2459
Gly	GGG	317263	0.2471
Total		1283759	
Stop	UAA	13963	
Stop	UAG	10631	
Stop	UGA	24607	

TABLE 3: Codon Usage Table for Mouse Genes (*Mus musculus*)

Amino Acid	Codon	Number	Frequency
Phe	UUU	150467	0.4321
Phe	UUC	197795	0.5679
Total		348262	
Leu	UUA	55635	0.0625
Leu	UUG	116210	0.1306
Leu	CUU	114699	0.1289
Leu	CUC	179248	0.2015
Leu	CUA	69237	0.0778
Leu	CUG	354743	0.3987
Total		889772	
Ile	AUU	137513	0.3367
Ile	AUC	208533	0.5106
Ile	AUA	62349	0.1527
Total		408395	
Met	AUG	204546	1.0000
Total		204546	
Val	GUU	93754	0.1673
Val	GUC	140762	0.2513
Val	GUA	64417	0.1150
Val	GUG	261308	0.4664
Total		560241	
Ser	UCU	139576	0.1936
Ser	UCC	160313	0.2224
Ser	UCA	100524	0.1394
Ser	UCG	38632	0.0536
Ser	AGU	108413	0.1504
Ser	AGC	173518	0.2407
Total		720976	
Pro	CCU	162613	0.3036
Pro	CCC	164796	0.3077
Pro	CCA	151091	0.2821
Pro	CCG	57032	0.1065
Total		535532	

Amino Acid	Codon	Number	Frequency
Thr	ACU	119832	0.2472
Thr	ACC	172415	0.3556
Thr	ACA	140420	0.2896
Thr	ACG	52142	0.1076
Total		484809	
Ala	GCU	178593	0.2905
Ala	GCC	236018	0.3839
Ala	GCA	139697	0.2272
Ala	GCG	60444	0.0983
Total		614752	
Tyr	UAU	108556	0.4219
Tyr	UAC	148772	0.5781
Total		257328	
His	CAU	88786	0.3973
His	CAC	134705	0.6027
Total		223491	
Gln	CAA	101783	0.2520
Gln	CAG	302064	0.7480
Total		403847	
Asn	AAU	138868	0.4254
Asn	AAC	187541	0.5746
Total		326409	
Lys	AAA	188707	0.3839
Lys	AAG	302799	0.6161
Total		491506	
Asp	GAU	189372	0.4414
Asp	GAC	239670	0.5586
Total		429042	
Glu	GAA	235842	0.4015
Glu	GAG	351582	0.5985
Total		587424	
Cys	UGU	97385	0.4716
Cys	UGC	109130	0.5284
Total		206515	

Amino Acid	Codon	Number	Frequency
Trp	UGG	112588	1.0000
Total		112588	
Arg	CGU	41703	0.0863
Arg	CGC	86351	0.1787
Arg	CGA	58928	0.1220
Arg	CGG	92277	0.1910
Arg	AGA	101029	0.2091
Arg	AGG	102859	0.2129
Total		483147	
Gly	GGU	103673	0.1750
Gly	GGC	198604	0.3352
Gly	GGA	151497	0.2557
Gly	GGG	138700	0.2341
Total		592474	
Stop	UAA	5499	
Stop	UAG	4661	
Stop	UGA	10356	

TABLE 4: Codon Usage Table for Domestic Cat Genes (*Felis cattus*)

Amino Acid	Codon	Number	Frequency of usage
Phe	UUU	1204.00	0.4039
Phe	UUC	1777.00	0.5961
Total		2981	
Leu	UUA	404.00	0.0570
Leu	UUG	857.00	0.1209
Leu	CUU	791.00	0.1116
Leu	CUC	1513.00	0.2135
Leu	CUA	488.00	0.0688
Leu	CUG	3035.00	0.4282
Total		7088	
Ile	AUU	1018.00	0.2984
Ile	AUC	1835.00	0.5380
Ile	AUA	558.00	0.1636
Total		3411	

Amino Acid	Codon	Number	Frequency of usage
Met	AUG	1553.00	0.0036
Total		1553	
Val	GUU	696.00	0.1512
Val	GUC	1279.00	0.2779
Val	GUA	463.00	0.1006
Val	GUG	2164.00	0.4702
Total		4602	
Ser	UCU	940.00	0.1875
Ser	UCC	1260.00	0.2513
Ser	UCA	608.00	0.1213
Ser	UCG	332.00	0.0662
Ser	AGU	672.00	0.1340
Ser	AGC	1202.00	0.2397
Total		5014	
Pro	CCU	958.00	0.2626
Pro	CCC	1375.00	0.3769
Pro	CCA	850.00	0.2330
Pro	CCG	465.00	0.1275
Total		3648	
Thr	ACU	822.00	0.2127
Thr	ACC	1574.00	0.4072
Thr	ACA	903.00	0.2336
Thr	ACG	566.00	0.1464
Total		3865	
Ala	GCU	1129.00	0.2496
Ala	GCC	1951.00	0.4313
Ala	GCA	883.00	0.1952
Ala	GCG	561.00	0.1240
Total		4524	
Tyr	UAU	837.00	0.3779
Tyr	UAC	1378.00	0.6221
Total		2215	
His	CAU	594.00	0.3738
His	CAC	995.00	0.6262
Total		1589	

Amino Acid	Codon	Number	Frequency of usage
Gln	CAA	747.00	0.2783
Gln	CAG	1937.00	0.7217
Total		2684	
Asn	AAU	1109.00	0.3949
Asn	AAC	1699.00	0.6051
Total		2808	
Lys	AAA	1445.00	0.4088
Lys	AAG	2090.00	0.5912
Total		3535	
Asp	GAU	1255.00	0.4055
Asp	GAC	1840.00	0.5945
Total		3095	
Glu	GAA	1637.00	0.4164
Glu	GAG	2294.00	0.5836
Total		3931	
Cys	UGU	719.00	0.4425
Cys	UGC	906.00	0.5575
Total		1625	
Trp	UGG	1073.00	1.0000
Total		1073	
Arg	CGU	236.00	0.0700
Arg	CGC	629.00	0.1865
Arg	CGA	354.00	0.1050
Arg	CGG	662.00	0.1963
Arg	AGA	712.00	0.2112
Arg	AGG	779.00	0.2310
Total		3372	
Gly	GGU	648.00	0.1498
Gly	GGC	1536.00	0.3551
Gly	GGA	1065.00	0.2462
Gly	GGG	1077.00	0.2490
Total		4326	

Amino Acid	Codon	Number	Frequency of usage
Stop	UAA	55	
Stop	UAG	36	
Stop	UGA	110	

TABLE 5: Codon Usage Table for Cow Genes (*Bos taurus*)

Amino Acid	Codon	Number	Frequency of usage
Phe	UUU	13002	0.4112
Phe	UUC	18614	0.5888
Total		31616	
Leu	UUA	4467	0.0590
Leu	UUG	9024	0.1192
Leu	CUU	9069	0.1198
Leu	CUC	16003	0.2114
Leu	CUA	4608	0.0609
Leu	CUG	32536	0.4298
Total		75707	
Ile	AUU	12474	0.3313
Ile	AUC	19800	0.5258
Ile	AUA	5381	0.1429
Total		37655	
Met	AUG	17770	1.0000
Total		17770	
Val	GUU	8212	0.1635
Val	GUC	12846	0.2558
Val	GUA	4932	0.0982
Val	GUG	24222	0.4824
Total		50212	
Ser	UCU	10287	0.1804
Ser	UCC	13258	0.2325
Ser	UCA	7678	0.1347
Ser	UCG	3470	0.0609
Ser	AGU	8040	0.1410
Ser	AGC	14279	0.2505
Total		57012	

Amino Acid	Codon	Number	Frequency of usage
Pro	CCU	11695	0.2684
Pro	CCC	15221	0.3493
Pro	CCA	11039	0.2533
Pro	CCG	5621	0.1290
Total		43576	
Thr	ACU	9372	0.2203
Thr	ACC	16574	0.3895
Thr	ACA	10892	0.2560
Thr	ACG	5712	0.1342
Total		42550	
Ala	GCU	13923	0.2592
Ala	GCC	23073	0.4295
Ala	GCA	10704	0.1992
Ala	GCG	6025	0.1121
Total		53725	
Tyr	UAU	9441	0.3882
Tyr	UAC	14882	0.6118
Total		24323	
His	CAU	6528	0.3649
His	CAC	11363	0.6351
Total		17891	
Gln	CAA	8060	0.2430
Gln	CAG	25108	0.7570
Total		33168	
Asn	AAU	12491	0.4088
Asn	AAC	18063	0.5912
Total		30554	
Lys	AAA	17244	0.3897
Lys	AAG	27000	0.6103
Total		44244	
Asp	GAU	16615	0.4239
Asp	GAC	22580	0.5761
Total		39195	

Amino Acid	Codon	Number	Frequency of usage
Glu	GAA	21102	0.4007
Glu	GAG	31555	0.5993
Total		52657	
Cys	UGU	7556	0.4200
Cys	UGC	10436	0.5800
Total		17992	
Trp	UGG	10706	1.0000
Total		10706	
Arg	CGU	3391	0.0824
Arg	CGC	7998	0.1943
Arg	CGA	4558	0.1108
Arg	CGG	8300	0.2017
Arg	AGA	8237	0.2001
Arg	AGG	8671	0.2107
Total		41155	
Gly	GGU	8508	0.1616
Gly	GGC	18517	0.3518
Gly	GGA	12838	0.2439
Gly	GGG	12772	0.2427
Total		52635	
Stop	UAA	555	
Stop	UAG	394	
Stop	UGA	392	

By utilizing these or similar tables, one of ordinary skill in the art can apply the frequencies to any given polypeptide sequence, and produce a nucleic acid fragment of a codon-optimized coding region which encodes the polypeptide, but which uses codons optimal for a given species. Codon-optimized coding regions can be designed by various different methods.

In one method, a codon usage table is used to find the single most frequent codon used for any given amino acid, and that codon is used each time that particular amino acid appears in the polypeptide sequence. For example, referring to Table 2 above, for leucine, the most frequent codon is

CUG, which is used 41% of the time. Thus all the leucine residues in a given amino acid sequence would be assigned the codon CUG. Human codon-optimized nucleotide sequences encoding native PA (GenBank Accession Number AAA2263 (SEQ ID NO:4)) and native LF (GenBank Accession Number AAA22569 (SEQ ID NO:12)) which have been optimized using this method are presented herein as SEQ ID NO:21 and SEQ ID NO:22, respectively.

In another method, the actual frequencies of the codons are distributed randomly throughout the coding sequence. Thus using this method for optimization, if a hypothetical polypeptide sequence had 100 leucine residues, referring to Table 2 for frequency of usage in the humans, about 7, or 7% of the leucine codons would be UUA, about 13, or 13% of the leucine codons would be UUG, about 13, or 13% of the leucine codons would be CUU, about 20, or 20% of the leucine codons would be CUC, about 7, or 7% of the leucine codons would be CUA, and about 41, or 41% of the leucine codons would be CUG. These frequencies would be distributed randomly throughout the leucine codons in the coding region encoding the hypothetical polypeptide. As will be understood by those of ordinary skill in the art, the distribution of codons in the sequence will can vary significantly using this method, however, the sequence always encodes the same polypeptide. Three different human codon-optimized nucleotide sequences encoding native PA (GenBank Accession Number AAA2263 (SEQ ID NO:4)) which have been optimized using this method are presented herein as SEQ ID NO:23, SEQ ID NO:24, and SEQ ID NO:25. Three different human codon-optimized sequences encoding native LF (GenBank Accession Number AAA22569 (SEQ ID NO:12)) which have been optimized using this method are presented herein as SEQ ID NO:21 and SEQ ID NO:22, respectively.

When using the latter method, the term "about" is used precisely to account for fractional percentages of codon frequencies for a given amino acid. As used herein, "about" is defined as one amino acid more or one amino acid less than the value given. The whole number value of amino acids is

rounded up if the fractional frequency of usage is 0.50 or greater, and is rounded down if the fractional frequency of use is 0.49 or less. Using again the example of the frequency of usage of leucine in human genes for a hypothetical polypeptide having 62 leucine residues, the fractional frequency of codon usage would be calculated by multiplying 62 by the frequencies for the various codons. Thus, 7.28 percent of 62 equals 4.51 UUA codons, or "about 5," *i.e.*, 4, 5, or 6 UUA codons, 12.66 percent of 62 equals 7.85 UUG codons or "about 8," *i.e.*, 7, 8, or 9 UUG codons, 12.87 percent of 62 equals 7.98 CUU codons, or "about 8," *i.e.*, 7, 8, or 9 CUU codons, 19.56 percent of 62 equals 12.13 CUC codons or "about 12," *i.e.*, 11, 12, or 13 CUC codons, 7.00 percent of 62 equals 4.34 CUA codons or "about 4," *i.e.*, 3, 4, or 5 CUA codons, and 40.62 percent of 62 equals 25.19 CUG codons, or "about 25," *i.e.*, 24, 25, or 26 CUG codons.

Randomly assigning codons at an optimized frequency to encode a given polypeptide sequence, can be done manually by calculating codon frequencies for each amino acid, and then assigning the codons to the polypeptide sequence randomly. Additionally, various algorithms and computer software programs are readily available to those of ordinary skill in the art. For example, the "EditSeq" function in the Lasergene Package, available from DNAstar, Inc., Madison, WI, the backtranslation function in the VectorNTI Suite, available from InforMax, Inc., Bethesda, MD, and the "backtranslate" function in the GCG--Wisconsin Package, available from Accelrys, Inc., San Diego, CA. In addition, various resources are publicly available to codon-optimize coding region sequences. For example, the "backtranslation" function at http://www.entelechon.com/eng/backtranslation.html (visited July 9, 2002), the "backtranseq" function available at http://bioinfo.pbi.nrc.ca:8090/EMBOSS/index.html (visited July 9, 2002). Constructing a rudimentary algorithm to assign codons based on a given frequency can also easily be accomplished with basic mathematical functions by one of ordinary skill.

A number of options are available for synthesizing codon optimized coding regions designed by any of the methods described above, using standard and routine molecular biological manipulations well known to those of ordinary skill in the art. In one approach, a series of complementary oligonucleotide pairs of 80-90 nucleotides each in length and spanning the length of the desired sequence are synthesized by standard methods. These oligonucleotide pairs are synthesized such that upon annealing, they form double stranded fragments of 80-90 base pairs, containing cohesive ends, *e.g.*, each oligonucleotide in the pair is synthesized to extend 3, 4, 5, 6, 7, 8, 9, 10, or more bases beyond the region that is complementary to the other oligonucleotide in the pair. The single-stranded ends of each pair of oligonucleotides is designed to anneal with the single-stranded end of another pair of oligonucleotides. The oligonucleotide pairs are allowed to anneal, and approximately five to six of these double-stranded fragments are then allowed to anneal together via the cohesive single stranded ends, and then they ligated together and cloned into a standard bacterial cloning vector, for example, a TOPO® vector available from Invitrogen Corporation, Carlsbad, CA. The construct is then sequenced by standard methods. Several of these constructs consisting of 5 to 6 fragments of 80 to 90 base pair fragments ligated together, *i.e.*, fragments of about 500 base pairs, are prepared, such that the entire desired sequence is represented in a series of plasmid constructs. The inserts of these plasmids are then cut with appropriate restriction enzymes and ligated together to form the final construct. The final construct is then cloned into a standard bacterial cloning vector, and sequenced. Additional methods would be immediately apparent to the skilled artisan. In addition, gene synthesis is readily available commercially.

In certain embodiments, an entire polypeptide sequence, or fragment, variant, or derivative thereof is codon optimized by any of the methods described herein. Various desired fragments, variants or derivatives are designed, and each is then codon-optimized individually. In addition, partially codon-optimized coding regions of the present invention can be designed and

constructed. For example, the invention includes a nucleic acid fragment of a codon-optimized coding region encoding a polypeptide in which at least about 1%, 2%, 3,% 4%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% of the codon positions have been codon-optimized for a given species. That is, they contain a codon that is preferentially used in the genes of a desired species, *e.g.*, a vertebrate species, *e.g.*, humans, in place of a codon that is normally used in the native nucleic acid sequence.

In additional embodiments, a full-length polypeptide sequence is codon-optimized for a given species resulting in a codon-optimized coding region encoding the entire polypeptide, and then nucleic acid fragments of the codon-optimized coding region, which encode fragments, variants, and derivatives of the polypeptide are made from the original codon-optimized coding region. As would be well understood by those of ordinary skill in the art, if codons have been randomly assigned to the full-length coding region based on their frequency of use in a given species, nucleic acid fragments encoding fragments, variants, and derivatives would not necessarily be *fully* codon optimized for the given species. However, such sequences are still much closer to the codon usage of the desired species than the native codon usage. The advantage of this approach is that synthesizing codon-optimized nucleic acid fragments encoding each fragment, variant, and derivative of a given polypeptide, although routine, would be time consuming and would result in significant expense.

The codon-optimized coding regions can be versions encoding any gene products from any strain of *Bacillus anthracis*, or fragments, variants, or derivatives of such gene products. Described herein are nucleic acid fragments of codon-optimized coding regions encoding the *Bacillus anthracis* protective antigen (PA) gene and the *Bacillus anthracis* lethal factor (LF), the nucleic acid fragments encoding the complete polypeptide, as well as various fragments, variants, and derivatives thereof, although other PA or LF - encoding nucleic acid sources are not excluded.

The present invention is directed to compositions and methods of enhancing the immune response of a vertebrate in need of protection against *Bacillus anthracis* infection by administering *in vivo*, into a tissue of a vertebrate, a polynucleotide comprising a codon-optimized coding region encoding a polypeptide of *Bacillus anthracis*, or a nucleic acid fragment of such a coding region encoding a fragment, variant or derivative thereof. Codon optimization is carried out for a particular vertebrate species by methods described herein, for example, in certain embodiments codon-optimized coding regions encoding polypeptides of *Bacillus anthracis*, or nucleic acid fragments of such coding regions encoding fragments, variants, or derivatives thereof are optimized according to human codon usage. The polynucleotides of the invention are incorporated into the cells of the vertebrate *in vivo*, and an immunologically effective amount of a *Bacillus anthracis* polypeptide is produced *in vivo*.In particular, the present invention relates to codon-optimized coding regions encoding polypeptides of *Bacillus anthracis*, or nucleic acid fragments of such coding regions fragments, variants, or derivatives thereof which have been optimized according to mammalian codon usage, for example, human codon usage, cow codon usage, domestic cat codon usage, or mouse codon usage. For example, human codon-optimized coding regions encoding polypeptides of *Bacillus anthracis*, or nucleic acid fragments of such coding regions encoding fragments, variants, or derivatives thereof are prepared by incorporating codons preferred for use in human genes into the DNA sequence encoding the *B. anthracis* polypeptide. Also provided are polynucleotides, vectors, and other expression constructs comprising codon-optimized coding regions encoding polypeptides of *Bacillus anthracis*, or nucleic acid fragments of such coding regions encoding fragments, variants, or derivatives thereof, and various methods of using such polynucleotides, vectors and other expression constructs.

The present invention is further directed towards polynucleotides comprising codon-optimized coding regions encoding polypeptides of *Bacillus anthracis* toxin, for example, *Bacillus anthracis* lethal toxin and its

component polypeptides, for example, lethal factor (LF) and protective antigen (PA). The invention is also directed to polynucleotides comprising codon-optimized nucleic acid fragments encoding fragments, variants and derivatives of these polypeptides.

The present invention provides isolated polynucleotides comprising codon-optimized coding regions of *Bacillus anthracis* PA, or fragments, variants, or derivatives thereof. In certain embodiments described herein, a codon-optimized coding region encoding SEQ ID NO:4 is optimized according to codon usage in humans (*Homo sapiens*). Alternatively, a codon-optimized coding region encoding SEQ ID NO:4 may be optimized according to codon usage in any plant, animal, or microbial species.

Codon-optimized coding regions encoding SEQ ID NO:4, optimized according to codon usage in humans are designed as follows. The amino acid composition of SEQ ID NO:4 is shown in Table 6.

TABLE 6

Amino Acid		Number in SEQ ID NO:4
A	Ala	41
R	Arg	29
C	Cys	0
G	Gly	36
H	His	10
I	Ile	57
L	Leu	62
K	Lys	60
M	Met	10
F	Phe	24
P	Pro	29
S	Ser	72
T	Thr	58
W	Trp	7
Y	Tyr	28
V	Val	43
N	Asn	69

D	Asp	47
Q	Gln	31
E	Glu	51

Using the amino acid composition shown in Table 6, a human codon-optimized coding region which encodes SEQ ID NO:4 can be designed by any of the methods discussed herein. In the first approach, each amino acid is assigned the most frequent codon used in the human genome for that amino acid. According to this method, codons are assigned to the coding region encoding SEQ ID NO:4 as follows: the 24 phenylalanine codons are TTC, the 62 leucine codons are CTG, the 57 isoleucine codons are ATC, the 10 methionine codons are ATG, the 43 valine codons are GTG, the 72 serine codons are AGC, the 29 proline codons are CCC, the 58 threonine codons are ACC, the 41 alanine codons are GCC, the 28 tyrosine codons are TAC, the 10 histidine codons are CAC, the 31 glutamine codons are CAG, the 69 asparagine codons are AAC, the 60 lysine codons are AAG, the 47 aspartic acid codons are GAC, the 51 glutamic acid codons are GAG, the 7 tryptophan codons are TGG, the 29 arginine codons are CGG, AGA, or AGG (the frequencies of usage of these three codons in the human genome are not significantly different), and the 36 glycine codons are GGC. The codon-optimized PA coding region designed by this method is presented herein as SEQ ID NO:21.

Alternatively, a human codon-optimized coding region which encodes SEQ ID NO:4 can be designed by randomly assigning each of any given amino acid a codon based on the frequency that codon is used in the human genome. These frequencies are shown in Table 2 above. Using this latter method, codons are assigned to the coding region encoding SEQ ID NO:4 as follows: about 11 of the 24 phenylalanine codons are TTT, and about 13 of the phenylalanine codons are TTC; about 5 of the 62 leucine codons are TTA, about 8 of the leucine codons are TTG, about 8 of the leucine codons are CTT, about 12 of the leucine codons are CTC, about 4 of the leucine codons are CTA, and about 25 of the leucine codons are CTG; about 20 of the 57

isoleucine codons are ATT, about 28 of the isoleucine codons are ATC, and about 9 of the isoleucine codons are ATA; the 10 methionine codons are ATG; about 8 of the 43 valine codons are GTT, about 10 of the valine codons are GTG, about 5 of the valine codons are GTA, and about 20 of the valine codons are GTG; about 13 of the 72 serine codons are TCT, about 16 of the serine codons are TCC, about 11 of the serine codons are TCA, about 4 of the serine codons are TCG, about 11 of the serine codons are AGT, and about 17 of the serine codons are AGC; about 8 of the 29 proline codons are CCT, about 10 of the proline codons are CCC, about 8 of the proline codons are CCA, and about 3 of the proline codons are CCG; about 14 of the 58 threonine codons are ACT, about 21 of the threonine codons are ACC, about 16 of the threonine codons are ACA, and about 7 of the threonine codons are ACG; about 11 of the 41 alanine codons are GGT, about 17 of the alanine codons are GCC, about 9 of the alanine codons are GCA, and about 4 of the alanine codons are GCG; about 12 of the 28 tyrosine codons are TAT and about 16 of the tyrosine codons are TAC; about 4 of the 10 histidine codons are CAT and about 6 of the histidine codons are CAC; about 8 of the 31 glutamine codons are CAA and about 23 of the glutamine codons are CAG; about 32 of the 69 asparagine codons are AAT and about 37 of the asparagine codons are AAC; about 25 of the 60 lysine codons are AAA and about 35 of the lysine codons are AAG; about 22 of the 47 aspartic acid codons are GAT and about 25 of the aspartic acid codons are GAC; about 21 of the 51 glutamic acid codons are GAA and about 30 of the glutamic acid codons are GAG; the 7 tryptophan codons are TGG; about 2 of the 29 arginine codons are CGT, about 6 of the arginine codons are CGC, about 3 of the arginine codons are CGA, about 6 of the arginine codons are CGG, about 6 of the arginine codons are AGA, and about 6 of the arginine codons are AGG; and about 6 of the 36 glycine codons are GGT, about 12 of the glycine codons are GGC, about 9 of the glycine codons are GGA, and about 9 of the glycine codons are GGG.

As described above, the term "about" means that the number of amino acids encoded by a certain codon may be one more or one less than the

number given. It would be understood by those of ordinary skill in the art that the total number of any amino acid in the polypeptide sequence must remain constant, therefore, if there is one "more" of one codon encoding a give amino acid, there would have to be one "less" of another codon encoding that same amino acid.

Representative codon-optimized coding regions encoding SEQ ID NO:4, optimized according to codon usage in humans designed by this method are presented herein as SEQ ID NO:23, SEQ ID NO:24, and SEQ ID NO:25.

In certain embodiments, the present invention provides an isolated polynucleotide comprising a nucleic acid fragment which encodes at least 10, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 95, or at least 100 or more contiguous amino acids of SEQ ID NO:4, where the nucleic acid fragment is a fragment of a codon-optimized coding region encoding SEQ ID NO:4. The codon optimized coding region can be optimized according to codon usage in any species, for example any vertebrate species, for example any mammalian species, for example human.

Further provided is an isolated polynucleotide comprising a nucleic acid fragment of a codon-optimized coding region encoding SEQ ID NO:4, where the nucleic acid fragment encodes amino acids 199 to 764 of SEQ ID NO:4. This polypeptide fragment is the 63-kD furin cleavage product (PA63) of the 82-kD protective antigen precursor polypeptide (PA83). The codon optimized coding region can be optimized according to codon usage in any species, for example any vertebrate species, for example any mammalian species, for example human. Included in this embodiment are nucleic acid fragments of a human codon-optimized coding region encoding SEQ ID NO:4, selected from: nucleotides 82 to 1779 of SEQ ID NO:1 (shown in Fig. 1), nucleotides 595 to 2292 of SEQ ID NO:23, nucleotides 595 to 2292 of SEQ ID NO:24, and nucleotides 595 to 2292 of SEQ ID NO:25.

Further provided is an isolated polynucleotide comprising a nucleic acid fragment of a codon-optimized coding region encoding SEQ ID NO:4,

where the nucleic acid fragment encodes amino acids 30 to 764 of SEQ ID NO:4. This polypeptide fragment is the mature full-length PA, *i.e.*, PA83. The codon optimized coding region can be optimized according to codon usage in any species, for example any vertebrate species, for example any mammalian species, for example human. Included in this embodiment are nucleic acid fragments of a human codon-optimized coding region encoding SEQ ID NO:4, selected from: nucleotides 88 to 2292 of SEQ ID NO:23, nucleotides 88 to 2292 of SEQ ID NO:24, and nucleotides 88 to 2292 of SEQ ID NO:25.

In certain embodiments, the present invention provides an isolated polynucleotide comprising a nucleic acid fragment which encodes a polypeptide at least 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to PA63, *i.e.*, amino acids 199 to 764 of SEQ ID NO:4, and where the nucleic acid fragment is a variant of a codon-optimized coding region encoding SEQ ID NO:4. The codon optimized coding region can be optimized according to codon usage in any species, for example any vertebrate species, for example any mammalian species, for example human.

Further provided is an isolated polynucleotide comprising a nucleic acid fragment which encodes a polypeptide variant of PA63, *i.e.*, amino acids 199 to 764 of SEQ ID NO:4, in which the amino acids corresponding to amino acids 342 and 343 of SEQ ID NO:4 have been deleted, and where the nucleic acid fragment is a variant of a codon-optimized coding region encoding SEQ ID NO:4. This variation in the amino acid sequence of PA63 eliminates two phenylalanine residues thought to be important in forming the pore in the *B. anthracis* lethal toxin. The codon optimized coding region can be optimized according to codon usage in any species, for example any vertebrate species, for example any mammalian species, for example human. Included in this embodiment is a nucleic acid fragment which is a variant of a human codon-optimized coding region encoding SEQ ID NO:4, where the nucleic acid fragment encodes amino acids 24 to 564 of SEQ ID NO:6 (shown in Fig. 2). Also included in this embodiment is a nucleic acid fragment comprising, or

alternatively consisting of nucleotides 82 to 1773 of SEQ ID NO:5 (shown in Fig. 2).

Further provided is an isolated polynucleotide comprising a nucleic acid fragment which encodes a polypeptide variant of PA63, *i.e.*, amino acids 199 to 764 of SEQ ID NO:4, in which the asparagine residues at positions corresponding to amino acids 275, 321, 357, 417, 505, 538, 599, 650, 693, and 738 of SEQ ID NO:4 have been each replaced with an amino acids other than asparagine, and where the nucleic acid fragment is a variant of a codon-optimized coding region encoding SEQ ID NO:4. In certain embodiments, the asparagine residues at positions corresponding to amino acids 275, 321, 357, 417, 505, 538, 599, 650, 693, and 738 of SEQ ID NO:4 have been each replaced with glutamine residues, where the nucleic acid fragment is a variant of a codon-optimized coding region encoding SEQ ID NO:4. Either of these variations in the amino acid sequence of PA63 removes adventitous substrates for asparagine-linked glycosylation present in the amino acid sequence. The codon optimized coding region can be optimized according to codon usage in any species, for example any vertebrate species, for example any mammalian species, for example human. Included in this embodiment is a nucleic acid fragment which is a variant of a human codon-optimized coding region encoding SEQ ID NO:4, where the nucleic acid fragment encodes amino acids 24 to 566 of SEQ ID NO:18 (shown in Fig. 7). Also included in this embodiment is a nucleic acid fragment comprising, or alternatively consisting of nucleotides 82 to 1779 of SEQ ID NO:17 (shown in Fig. 7).

Further provided is an isolated polynucleotide comprising a nucleic acid fragment which encodes a polypeptide variant of PA63, *i.e.*, amino acids 199 to 764 of SEQ ID NO:4, in which the amino acids corresponding to amino acids 342 and 343 of SEQ ID NO:4 have been deleted, where the asparagine residues at positions corresponding to amino acids 275, 321, 357, 417, 505, 538, 599, 650, 693, and 738 of SEQ ID NO:4 have been each replaced with an amino acids other than asparagine, for example, glutamine, and where the nucleic acid fragment is a variant of a codon-optimized coding region

encoding SEQ ID NO:4. The codon optimized coding region can be optimized according to codon usage in any species, for example any vertebrate species, for example any mammalian species, for example human.

In certain embodiments, the present invention provides an isolated polynucleotide comprising a nucleic acid fragment which encodes a polypeptide at least 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to PA83, *i.e.*, amino acids 30 to 764 of SEQ ID NO:4, and where the nucleic acid fragment is a variant of a codon-optimized coding region encoding SEQ ID NO:4. The codon optimized coding region can be optimized according to codon usage in any species, for example any vertebrate species, for example any mammalian species, for example human.

Further provided is an isolated polynucleotide comprising a nucleic acid fragment which encodes a polypeptide variant of PA83, *i.e.*, amino acids 30 to 764 of SEQ ID NO:4, in which the amino acids corresponding to amino acids 192 to 197 of SEQ ID NO:4 have been deleted, and where the nucleic acid fragment is a variant of a codon-optimized coding region encoding SEQ ID NO:4. This variation in the amino acid sequence of PA83 eliminates the furin cleavage site in PA83, and thus the encoded polypeptide cannot be cleaved as a substrate for furin, and cannot form the pore of the lethal toxin of *B. anthracis*. The codon optimized coding region can be optimized according to codon usage in any species, for example any vertebrate species, for example any mammalian species, for example human. Included in this embodiment is a nucleic acid fragment which is a variant of a human codon-optimized coding region encoding SEQ ID NO:4, where the nucleic acid fragment encodes amino acids 24 to 752 of SEQ ID NO:8 (shown in Fig. 3). Also included in this embodiment is a nucleic acid fragment comprising, or alternatively consisting of nucleotides 82 to 2268 of SEQ ID NO:7 (shown in Fig. 3).

Further provided is an isolated polynucleotide comprising a nucleic acid fragment which encodes a polypeptide variant of PA83, *i.e.*, amino acids 30 to 764 of SEQ ID NO:4, in which the asparagine residues at positions

corresponding to amino acids 39, 153, 275, 321, 357, 417, 505, 538, 599, 650, 693, and 738 of SEQ ID NO:4 have been each replaced with an amino acids other than asparagine, and where the nucleic acid fragment is a variant of a codon-optimized coding region encoding SEQ ID NO:4. In certain embodiments, the asparagine residues at positions corresponding to amino acids 39, 153, 275, 321, 357, 417, 505, 538, 599, 650, 693, and 738 of SEQ ID NO:4 have been each replaced with glutamine residues, where the nucleic acid fragment is a variant of a codon-optimized coding region encoding SEQ ID NO:4. Either of these variations in the amino acid sequence of PA83 removes adventitous substrates for asparagine-linked glycosylation present in the amino acid sequence. The codon optimized coding region can be optimized according to codon usage in any species, for example any vertebrate species, for example any mammalian species, for example human.

Further provided is an isolated polynucleotide comprising a nucleic acid fragment which encodes a polypeptide variant of PA83, *i.e.*, amino acids 30 to 764 of SEQ ID NO:4, in which the amino acids corresponding to amino acids 192 to 197 of SEQ ID NO:4 have been deleted, where the asparagine residues at positions corresponding to amino acids 39, 153, 275, 321, 357, 417, 505, 538, 599, 650, 693, and 738 of SEQ ID NO:4 have been each replaced with an amino acids other than asparagine, for example, glutamine, and where the nucleic acid fragment is a variant of a codon-optimized coding region encoding SEQ ID NO:4. The codon optimized coding region can be optimized according to codon usage in any species, for example any vertebrate species, for example any mammalian species, for example human.

The present invention provides isolated polynucleotides comprising codon-optimized coding regions of *Bacillus anthracis* LF, or fragments, variants, or derivatives thereof. In certain embodiments described herein, a codon-optimized coding region encoding SEQ ID NO:12 is optimized according to codon usage in humans (*Homo sapiens*). Alternatively, a codon-optimized coding region encoding SEQ ID NO:12 may be optimized according to codon usage in any plant, animal, or microbial species.

Codon-optimized coding regions encoding SEQ ID NO:12, optimized according to codon usage in humans are designed as follows. The amino acid composition of SEQ ID NO:12 is shown in Table 7.

TABLE 7

Amino Acid		Number in SEQ ID NO:12
A	Ala	34
R	Arg	27
C	Cys	1
G	Gly	35
H	His	21
I	Ile	74
L	Leu	80
K	Lys	86
M	Met	10
F	Phe	29
P	Pro	21
S	Ser	54
T	Thr	28
W	Trp	5
Y	Tyr	35
V	Val	40
N	Asn	54
D	Asp	55
Q	Gln	41
E	Glu	79

Using the amino acid composition shown in Table 7, a human codon-optimized coding region which encodes SEQ ID NO:12 can be designed by any of the methods discussed herein. In the first approach, each amino acid is assigned the most frequent codon used in the human genome for that amino acid. According to this method, codons are assigned to the coding region encoding SEQ ID NO:4 as follows: the 29 phenylalanine codons are TTC, the 80 leucine codons are CTG, the 74 isoleucine codons are ATC, the 10 methionine codons are ATG, the 43 valine codons are GTG, the 54 serine codons are AGC, the 21 proline codons are CCC, the 28 threonine codons are

ACC, the 34 alanine codons are GCC, the 35 tyrosine codons are TAC, the 21 histidine codons are CAC, the 41 glutamine codons are CAG, the 54 asparagine codons are AAC, the 86 lysine codons are AAG, the 55 aspartic acid codons are GAC, the 79 glutamic acid codons are GAG, the 5 tryptophan codons are TGG, the 27 arginine codons are CGG, AGA, or AGG (the frequencies of usage of these three codons in the human genome are not significantly different), and the 35 glycine codons are GGC. The codon-optimized LF coding region designed by this method is presented herein as SEQ ID NO:22.

Alternatively, a human codon-optimized coding region which encodes SEQ ID NO:12 can be designed by randomly assigning each of any given amino acid a codon based on the frequency that codon is used in the human genome. These frequencies are shown in Table 2 above. Using this latter method, codons are assigned to the coding region encoding SEQ ID NO:12 as follows: about 13 of the 29 phenylalanine codons are TTT and about 16 of the phenylalanine codons are TTC; about 6 of the 80 leucine codons are TTA, about 10 of the leucine codons are TTG, about 10 of the leucine codons are CTT, about 16 of the leucine codons are CTC, about 6 of the leucine codons are CTA, and about 32 of the leucine codons are CTG; about 26 of the 74 isoleucine codons are ATT, about 36 of the isoleucine codons are ATC, and about 12 of the isoleucine codons are ATA; the 10 methionine codons are ATG; about 7 of the 40 valine codons are GTT, about 9 of the valine codons are GTG, about 5 of the valine codons are GTA, and about 19 of the valine codons are GTG; about 10 of the 54 serine codons are TCT, about 12 of the serine codons are TCC, about 8 of the serine codons are TCA, about 3 of the serine codons are TCG, about 8 of the serine codons are AGT, and about 13 of the serine codons are AGC; about 6 of the 21 proline codons are CCT, about 7 of the proline codons are CCC, about 6 of the proline codons are CCA, and about 2 of the proline codons are CCG; about 7 of the 28 threonine codons are ACT, about 10 of the threonine codons are ACC, about 8 of the threonine codons are ACA, and about 3 of the threonine codons are ACG; about 9 of the

34 alanine codons are GGT, about 14 of the alanine codons are GCC, about 8 of the alanine codons are GCA, and about 3 of the alanine codons are GCG; about 15 of the 35 tyrosine codons are TAT and about 20 of the tyrosine codons are TAC; about 9 of the 21 histidine codons are CAT and about 12 of the histidine codons are CAC; about 10 of the 41 glutamine codons are CAA and about 31 of the glutamine codons are CAG; about 25 of the 54 asparagine codons are AAT and about 29 of the asparagine codons are AAC; about 36 of the 86 lysine codons are AAA and about 50 of the lysine codons are AAG; about 25 of the 55 aspartic acid codons are GAT and about 30 of the aspartic acid codons are GAC; about 33 of the 79 glutamic acid codons are GAA and about 46 of the glutamic acid codons are GAG; the single cysteine codon is either TGT or TGC; the 5 tryptophan codons are TGG; about 2 of the 27 arginine codons are CGT, about 5 of the arginine codons are CGC, about 3 of the arginine codons are CGA, about 6 of the arginine codons are CGG, about 6 of the arginine codons are AGA, and about 5 of the arginine codons are AGG; and about 6 of the 35 glycine codons are GGT, about 12 of the glycine codons are GGC, about 8 of the glycine codons are GGA, and about 9 of the glycine codons are GGG.

As described above, the term "about" means that the number of amino acids encoded by a certain codon may be one more or one less than the number given. It would be understood by those of ordinary skill in the art that the total number of any amino acid in the polypeptide sequence must remain constant, therefore, if there is one "more" of one codon encoding a give amino acid, there would have to be one "less" of another codon encoding that same amino acid.

Representative codon-optimized coding regions encoding SEQ ID NO:12, optimized according to codon usage in humans designed by this method are presented herein as SEQ ID NO:26, SEQ ID NO:27, and SEQ ID NO:28.

In certain embodiments, the present invention provides an isolated polynucleotide comprising a nucleic acid fragment which encodes at least 10,

at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 95, or at least 100 or more contiguous amino acids of SEQ ID NO:12, where the nucleic acid fragment is a fragment of a codon-optimized coding region encoding SEQ ID NO:12. The codon optimized coding region can be optimized according to codon usage in any species, for example any vertebrate species, for example any mammalian species, for example human.

Further provided is an isolated polynucleotide comprising a nucleic acid fragment of a codon-optimized coding region encoding SEQ ID NO:12, where the nucleic acid fragment encodes amino acids 34 to 809 of SEQ ID NO:12. This polypeptide fragment is the mature form of *B. anthracis* LF. The codon optimized coding region can be optimized according to codon usage in any species, for example any vertebrate species, for example any mammalian species, for example human. Included in this embodiment are nucleic acid fragments of a human codon-optimized coding region encoding amino acids 34 to 809 of SEQ ID NO:12, selected from: nucleotides 100 to 2427 of SEQ ID NO:26, nucleotides 100 to 2427 of SEQ ID NO:27, and nucleotides 100 to 2427 of SEQ ID NO:28.

Further provided is an isolated polynucleotide comprising a nucleic acid fragment of a codon-optimized coding region encoding SEQ ID NO:12, where the nucleic acid fragment encodes amino acids 34 to 583 of SEQ ID NO:12. This polypeptide fragment encodes domains I-III of mature *B. anthracis* LF, but not domain IV, the protease domain. The codon optimized coding region can be optimized according to codon usage in any species, for example any vertebrate species, for example any mammalian species, for example human. Included in this embodiment are nucleic acid fragments of a human codon-optimized coding region encoding SEQ ID NO:12, selected from: nucleotides 82 to 1731 of SEQ ID NO:13 (shown in Fig. 5), nucleotides 100 to 1752 of SEQ ID NO:26, nucleotides 100 to 1752 of SEQ ID NO:27, and nucleotides 100 to 1752 of SEQ ID NO:28.

Further provided is an isolated polynucleotide comprising a nucleic acid fragment of a codon-optimized coding region encoding SEQ ID NO:12, where the nucleic acid fragment encodes amino acids 34 to 254 of SEQ ID NO:12. This polypeptide fragment encodes a portion of domain I of mature *B. anthracis* LF, that directly binds to PA63. The codon optimized coding region can be optimized according to codon usage in any species, for example any vertebrate species, for example any mammalian species, for example human. Included in this embodiment are nucleic acid fragments of a human codon-optimized coding region encoding SEQ ID NO:12, selected from: nucleotides 82 to 744 of SEQ ID NO:15 (shown in Fig. 6), nucleotides 100 to 762 of SEQ ID NO:26, nucleotides 100 to 762 of SEQ ID NO:27, and nucleotides 100 to 762 of SEQ ID NO:28.

In certain embodiments, the present invention provides an isolated polynucleotide comprising a nucleic acid fragment which encodes a polypeptide at least 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to LF, *i.e.*, amino acids 34 to 809 of SEQ ID NO:12, and where the nucleic acid fragment is a variant of a codon-optimized coding region encoding SEQ ID NO:12. The codon optimized coding region can be optimized according to codon usage in any species, for example any vertebrate species, for example any mammalian species, for example human.

Further provided is an isolated polynucleotide comprising a nucleic acid fragment which encodes a polypeptide variant of LF, *i.e.*, amino acids 34 to 809 of SEQ ID NO:12, in which the histidine residues corresponding to amino acids 719 and 723 of SEQ ID NO:12 have been deleted, and replaced with an amino acid other than histidine, and/or the glutamic acid residue corresponding to amino acid 720 of SEQ ID NO:12 has been deleted and replaced with an amino acid other than glutamic acid, where the nucleic acid fragment is a variant of a codon-optimized coding region encoding SEQ ID NO:12. In certain embodiments, the histidine residues corresponding to amino acids 719 and 723 of SEQ ID NO:12 have been deleted, and replaced with

alanine residues, and/or the glutamic acid residue corresponding to amino acid 720 of SEQ ID NO:12 has been deleted and replaced with an aspartic acid residue, where the nucleic acid fragment is a variant of a codon-optimized coding region encoding SEQ ID NO:12. Any of these variations in the amino acid sequence of LF, either alone or in combination, eliminate the protease activity of LF. The codon optimized coding region can be optimized according to codon usage in any species, for example any vertebrate species, for example any mammalian species, for example human. Included in this embodiment is a nucleic acid fragment which is a variant of a human codon-optimized coding region encoding SEQ ID NO:12, where the nucleic acid fragment encodes amino acids 24 to 799 of SEQ ID NO:10 (shown in Fig. 4). Also included in this embodiment is a nucleic acid fragment comprising, or alternatively consisting of nucleotides 82 to 2409 of SEQ ID NO:9 (shown in Fig. 4).

Further provided is an isolated polynucleotide comprising a nucleic acid fragment which encodes a polypeptide variant of LF, *i.e.*, amino acids 34 to 809 of SEQ ID NO:12, in which the asparagine residues at positions corresponding to amino acids 62, 212, 286, 478, 712, 736, and 757 of SEQ ID NO:12 have been each replaced with an amino acids other than asparagine, and where the nucleic acid fragment is a variant of a codon-optimized coding region encoding SEQ ID NO:12. In certain embodiments, the asparagine residues at positions corresponding to amino acids 62, 212, 286, 478, 712, 736, and 757 of SEQ ID NO:12 have been each replaced with glutamine residues, where the nucleic acid fragment is a variant of a codon-optimized coding region encoding SEQ ID NO:12. Either of these variations in the amino acid sequence of LF remove adventitous substrates for asparagine-linked glycosylation present in the amino acid sequence. The codon optimized coding region can be optimized according to codon usage in any species, for example any vertebrate species, for example any mammalian species, for example human.

Further provided is an isolated polynucleotide comprising a nucleic acid fragment which encodes a polypeptide variant of LF, *i.e.*, amino acids 34 to 809 of SEQ ID NO:12, in which the histidine residues corresponding to amino acids 719 and 723 of SEQ ID NO:12 have been deleted, and replaced with an amino acid other than histidine, and/or the glutamic acid residue corresponding to amino acid 720 of SEQ ID NO:12 has been deleted and replaced with an amino acid other than glutamic acid, and the asparagine residues at positions corresponding to amino acids 62, 212, 286, 478, 712, 736, and 757 of SEQ ID NO:12 have been each replaced with an amino acids other than asparagine, and where the nucleic acid fragment is a variant of a codon-optimized coding region encoding SEQ ID NO:12. In certain embodiments, the histidine residues corresponding to amino acids 719 and 723 of SEQ ID NO:12 have been deleted, and replaced with alanine residues, and/or the glutamic acid residue corresponding to amino acid 720 of SEQ ID NO:12 has been deleted and replaced with an aspartic acid residue, and the asparagine residues at positions corresponding to amino acids 62, 212, 286, 478, 712, 736, and 757 of SEQ ID NO:12 have been each replaced with glutamine residues, where the nucleic acid fragment is a variant of a codon-optimized coding region encoding SEQ ID NO:12. Any of these variations in the amino acid sequence of LF, either alone or in combination, eliminate the protease activity of LF, and also, adventitous substrates for asparagine-linked glycosylation present in the amino acid sequence have been removed. The codon optimized coding region can be optimized according to codon usage in any species, for example any vertebrate species, for example any mammalian species, for example human. Included in this embodiment is a nucleic acid fragment which is a variant of a human codon-optimized coding region encoding SEQ ID NO:12, where the nucleic acid fragment encodes amino acids 24 to 799 of SEQ ID NO:20 (shown in Fig. 8). Also included in this embodiment is a nucleic acid fragment comprising, or alternatively consisting of nucleotides 82 to 2409 of SEQ ID NO:19 (shown in Fig. 8).

In this manner, the present invention provides a method of enhancing the level of polypeptide expression from delivered polynucleotides *in vivo* and/or facilitating uptake of the polynucleotides by the cells of a desired species, for example a vertebrate species, for example a mammalian species, for example humans. Accordingly, the present invention provides a method of treatment and prevention against *Bacillus anthracis* infection.

Methods and Administration

The present invention further provides methods for delivering a polypeptide into a vertebrate, which comprise administering to a vertebrate one or more of the compositions described herein; such that upon administration of compositions such as those described herein, a *B. anthracis* polypeptide is expressed in the vertebrate, in an amount sufficient generate an immune response to *B. anthracis*.

The term "vertebrate" is intended to encompass a singular "vertebrate" as well as plural "vertebrates," and comprises mammals and birds, as well as fish, reptiles, and amphibians.

The term "mammal" is intended to encompass a singular "mammal" and plural "mammals," and includes, but is not limited to humans; primates such as apes, monkeys, orangutans, and chimpanzees; canids such as dogs and wolves; felids such as cats, lions, and tigers; equids such as horses, donkeys, and zebras, food animals such as cows, pigs, and sheep; ungalates such as deer and giraffes; and ursids such as bears. In particular, the mammal can be a human subject, a food animal or a companion animal.

The present invention further provides a method for generating, enhancing or modulating an immune response to *B. anthracis* comprising administering to a vertebrate one or more of the compositions described herein. In this method, the composition includes an isolated polynucleotide comprising a human codon-optimized coding region encoding a polypeptide of *Bacillus anthracis*, or a nucleic acid fragment of such a coding region

encoding a fragment, variant, or derivative thereof. The polynucleotides are incorporated into the cells of the vertebrate *in vivo*, and an antigenic amount of the *Bacillus anthracis* polypeptide, or fragment, variant, or derivative thereof, is produced *in vivo*. Upon administration of the composition according to this method, the *Bacillus anthracis* polypeptide is expressed in the vertebrate in an amount sufficient to elicit an immune response. Such an immune response might be used, for example, to generate antibodies to *B. anthracis* for use in diagnostic assays or as laboratory reagents.

The present invention further provides a method for generating, enhancing, or modulating a protective and/or therapeutic immune response to *B. anthracis* in a vertebrate, comprising administering to a vertebrate in need of therapeutic and/or preventative immunity one or more of the compositions described herein. In this method, the composition includes an isolated polynucleotide comprising a human codon-optimized coding region encoding a polypeptide of *Bacillus anthracis*, or a nucleic acid fragment of such a coding region encoding a fragment, variant, or derivative thereof. The polynucleotides are incorporated into the cells of the vertebrate *in vivo*, and an immunologically effective amount of the *Bacillus anthracis* polypeptide, or fragment or variant is produced *in vivo*. Upon administration of the composition according to this method, the *Bacillus anthracis* polypeptide is expressed in the vertebrate in a therapeutically or prophylactically effective amount.

As used herein, an "immune response" refers to the ability of a vertebrate to elicit an immune reaction to a composition delivered to that vertebrate. Examples of immune responses include an antibody response or a cellular, *e.g.*, T-cell, response. One or more compositions of the present invention may be used to treat a vertebrate prophylactically, *e.g.*, as a prophylactic vaccine, to establish or enhance immunity to *B. anthracis* in a healthy vertebrate prior to exposure to *B. anthracis* or contraction of anthrax disease, thus preventing the disease or reducing the severity of disease symptoms. One or more compositions of the present invention may also be

used to treat a vertebrate already exposed to *B. anthracis*, or already suffering from anthrax disease to further stimulate the immune system of the vertebrate, thus reducing or eliminating the symptoms associated with that disease or disorder. As defined herein, "treatment of a vertebrate" refers to the use of One or more compositions of the present invention to prevent, cure, retard, or reduce the severity of anthrax disease symptoms in a vertebrate, and/or result in no worsening of anthrax disease over a specified period of time. It is not required that any composition of the present invention provide total immunity to *B. anthracis* or totally cure or eliminate all anthrax disease symptoms. As used herein, a "a vertebrate in need of therapeutic and/or preventative immunity" refers to a vertebrate which it is desirable to treat, *i.e.*, to prevent, cure, retard, or reduce the severity of anthrax disease symptoms, and/or result in no worsening of anthrax disease over a specified period of time.

In other embodiments, one or more compositions of the present invention are utilized in a "prime boost" regimen. In these embodiments, one or more polynucleotide vaccine compositions of the present invention are delivered to a vertebrate, thereby priming the immune response of the vertebrate to *B. anthracis*, and then a second immunogenic composition is utilized as a boost vaccination. One or more polynucleotide vaccine compositions of the present invention are used to prime immunity, and then a second immunogenic composition, *e.g.*, a recombinant viral vaccine or vaccines, a different polynucleotide vaccine, one or more purified subunit *Bacillus anthracis* proteins, *e.g.*, PA or LF or a variant, fragment, or derivative thereof, or the existing AVA anthrax vaccine, is used to boost the anti-*Bacillus anthracis* immune response. The polynucleotide vaccine compositions may comprise one or more vectors for expression of one or more *Bacillus anthracis* lethal toxin genes as described herein. In addition, polynucleotide prime vaccine and the later boost vaccine elicit an immune response to the same or similar antigens, or they may be to different antigens.

In another embodiment, vectors are prepared for expression in the recombinant virus vaccine and in transfected mammalian cells as part of a polynucleotide vaccine.

The terms "priming" or "primary" and "boost" or "boosting" are used herein to refer to the initial and subsequent immunizations, respectively, *i.e.*, in accordance with the definitions these terms normally have in immunology.

Sterile immunity is defined herein as the ability to completely inhibit the germination of anthrax spores into bacteria. If germination occurs, the bacteria produce Letx and surviving rabbits immunized against the PA antigen would be expected to generate a response to LF. Likewise, rabbits immunized with LF should have a measurable response to PA.

Antibodies induced by recombinant PA or by the commercial anthrax vaccine, AVA, have been shown to have potential activities other than neutralization, that may affect the outcome of an infection by anthrax. Among these potential activities is the effect of preventing germination of bacteria from the spores. (Welkos, S. *et al. Microbiology. 147*: 1677-85 (2001)). DNA vaccination may induce levels of antibody consistent with those that prevent germination. The absence of an increase in LF, PA, or neutralization titers, following infection, has been observed in animals vaccinated with DNA vaccines. This is in contrast to animals vaccinated twice with a commercial anthrax vaccine, AVA. While not being bound by theory, the DNA vaccine may induce antibodies that possess novel protective activities independent of lethal toxin neutralization.

In certain embodiments, one or more compositions of the present invention are delivered to a vertebrate by methods described herein, thereby achieving an effective immune response, and or an effective therapeutic or preventative immune response.

More specifically, the compositions of the present invention may be administered to any tissue of a vertebrate, including, but not limited to, muscle, skin, brain tissue, lung tissue, liver tissue, spleen tissue, bone marrow tissue, thymus tissue, heart tissue, *e.g.*, myocardium, endocardium, and

pericardium, lymph tissue, blood tissue, bone tissue, pancreas tissue, kidney tissue, gall bladder tissue, stomach tissue, intestinal tissue, testicular tissue, ovarian tissue, uterine tissue, vaginal tissue, rectal tissue, nervous system tissue, eye tissue, glandular tissue, tongue tissue, and connective tissue, *e.g.*, cartilage.

Furthermore, the compositions of the present invention may be administered to any internal cavity of a vertebrate, including, but not limited to, the lungs, the mouth, the nasal cavity, the stomach, the peritoneal cavity, the intestine, any heart chamber, veins, arteries, capillaries, lymphatic cavities, the uterine cavity, the vaginal cavity, the rectal cavity, joint cavities, ventricles in brain, spinal canal in spinal cord, the ocular cavities, the lumen of a duct of a salivary gland or a liver. When the compositions of the present invention is administered to the lumen of a duct of a salivary gland or a liver, the desired polypeptide is encoded in each of the salivary gland and the liver such that the polypeptide is delivered into the blood stream of the vertebrate from each of the salivary gland and the liver. Certain modes for administration to secretory organs of a gastrointestinal system using the salivary gland, liver and pancreas to release a desired polypeptide into the bloodstream is disclosed in U.S. Patent Nos. 5,837,693 and 6,004,944, both of which are incorporated herein by reference in their entireties.

In one embodiment, the compositions are administered to muscle, either skeletal muscle or cardiac muscle, or lung tissue. Specific, but non-limiting modes for administration to lung tissue are disclosed in Wheeler, C.J., *et al., Proc. Natl. Acad. Sci. USA 93*:11454-11459 (1996), which is incorporated herein by reference in its entirety.

According to the disclosed methods, compositions of the present invention can be administered by intramuscular (i.m.), subcutaneous (s.c.), or intrapulmonary routes. Other suitable routes of administration include, but not limited to intratracheal, transdermal, intraocular, intranasal, inhalation, intracavity, intravenous (i.v.), intraductal (*e.g.*, into the pancreas) and intraparenchymal (*i.e.*, into any tissue) administration. Transdermal delivery

includes, but not limited to intradermal (*e.g.*, into the dermis or epidermis), transdermal (*e.g.*, percutaneous) and transmucosal administration (*i.e.*, into or through skin or mucosal tissue). Intracavity administration includes, but not limited to adminitration into oral, vaginal, rectal, nasal, peritoneal, or intestinal cavities as well as, intrathecal (*i.e.*, into spinal canal), intraventricular (*i.e.*, into the brain ventricles or the heart ventricles), inraatrial (*i.e.*, into the heart atrium) and sub arachnoid (*i.e.*, into the sub arachnoid spaces of the brain) administration.

Any mode of administration can be used so long as the mode results in the expression of the desired peptide or protein, in the desired tissue, in an amount sufficient to generate an immune response to *B. anthracis* and/or to generate a prophylactically or therapeutically effective immune response to *B. anthracis* in a vertebrate in need of such response. Administration means of the present invention include needle injection, catheter infusion, biolistic injectors, particle accelerators (*e.g.*, "gene guns" or pneumatic "needleless" injectors) Med-E-Jet (Vahlsing, H., *et al., J. Immunol. Methods* 171,11-22 (1994)), Pigjet (Schrijver, R., *et al., Vaccine* 15, 1908-1916 (1997)), Biojector (Davis, H., *et al., Vaccine* 12, 1503-1509 (1994); Gramzinski, R., *et al., Mol. Med.* 4, 109-118 (1998)), AdvantaJet (Linmayer, I., *et al., Diabetes Care 9*:294-297 (1986)), Medi-jector (Martins, J., and Roedl, E. *J. Occup. Med. 21*:821-824 (1979)), gelfoam sponge depots, other commercially available depot materials (*e.g.*, hydrogels), osmotic pumps (*e.g.*, Alza minipumps), oral or suppositorial solid (tablet or pill) pharmaceutical formulations, topical skin creams, and decanting, use of polynucleotide coated suture (Qin, Y., *et al., Life Sciences* 65, 2193-2203 (1999)) or topical applications during surgery. Certain modes of administration are intramuscular needle-based injection and pulmonary application via catheter infusion. Each of the references cited in this paragraph is incorporated herein by reference in its entirety.

Determining an effective amount of one or more compositions of the present invention depends upon a number of factors including, for example, the antigen being expressed, *e.g.*, PA or LF or fragments, variants, or

derivatives thereof, the age and weight of the subject, the precise condition requiring treatment and its severity, and the route of administration. Based on the above factors, determining the precise amount, number of doses, and timing of doses are within the ordinary skill in the art and will be readily determined by the attending physician or veterinarian.

Compositions of the present invention may include various salts, excipients, delivery vehicles and/or auxilliary agents as are disclosed, *e.g.*, in U.S. Patent Application Publication 2002/0019358, published February 14, 2002, which is incorporated herein by reference in its entirety.

Furthermore, compositions of the present invention may include one or more transfection facilitating compounds that facilitate delivery of polynucleotides to the interior of a cell, and/or to a desired location within a cell. As used herein, the terms "transfection faciliating compound," "transfection facilitating agent," and "transfection faciliating material" are synonymous, and may be used interchangeably. It should be noted that certain transfection facilitating compounds may also be "adjuvants" as described *infra*, *i.e.*, in addition to facilitating delivery of polynucleotides to the interior of a cell, the compound acts to alter or increase the immune response to the antigen encoded by that polynucleotide. Examples of the transfection facilitating compounds include, but are not limited to inorganic materials such as calcium phosphate, alum (aluminum sulfate), and gold particles (*e.g.*, "powder" type delivery vehicles); peptides that are, for example, cationic, intercell targeting (for selective delivery to certain cell types), intracell targeting (for nucleor localization or endosomal escape), and ampipathic (helix forming or pore forming); proteins that are, for example, basic (*e.g.*, positively charged) such as histones, targeting (*e.g.*, asialoprotein), viral (*e.g.*, Sendai virus coat protein), and pore-forming; lipids that are, for example, cationic (*e.g.*, DMRIE, DOSPA, DC-Chol), basic (*e.g.*, steryl amine), neutral (*e.g.*, cholesterol), anionic (*e.g.*, phosphatidyl serine), and zwitterionic (*e.g.*, DOPE, DOPC); and polymers such as dendrimers, star-polymers, "homogenous" poly-amino acids (*e.g.*, poly-lysine, poly-arginine), "heterogenous"

poly-amino acids (*e.g.*, mixtures of lysine & glycine), co-polymers, polyvinylpyrrolidinone (PVP), and polyethylene glycol (PEG). A transfection facilitating material can be used alone or in combination with one or more other transfection facilitating materials. Two or more transfection facilitating materials can be combined by chemical bonding (*e.g.*, covalent and ionic such as in lipidated polylysine, PEGylated polylysine) (Toncheva, *et al.*, *Biochim. Biophys. Acta 1380(3)*:354-368 (1988)), mechical mixing (*e.g.*, free moving materials in liquid or solid phase such as "polylysine + cationic lipids") (Gao and Huang, *Biochemistry 35*:1027-1036 (1996); Trubetskoy, *et al., Biochem. Biophys. Acta 1131*:311-313 (1992)), and aggregation (*e.g.*, co-precipitation, gel forming such as in cationic lipids + poly-lactide co-galactide, and polylysine + gelatin).

One category of transfection facilitating materials is cationic lipids. Examples of cationic lipids are 5-carboxyspermylglycine dioctadecylamide (DOGS) and dipalmitoyl-phophatidylethanolamine-5carboxyspermylamide (DPPES). Cationic cholesterol derivatives are also useful, including {3β-[N-N',N'-dimethylamino)ethane]-carbomoyl}-cholesterol (DC-Chol). Dimethyldioctdecyl-ammonium bromide (DDAB), N-(3-aminopropyl)-N,N-(*bis*-(2-tetradecyloxyethyl))-N-methyl-ammonium bromide (PA-DEMO), N-(3-aminopropyl)-N,N-(*bis*-(2-dodecyloxyethyl))-N-methyl-ammonium bromide (PA-DELO), N,N,N-*tris*-(2-dodecyloxy)ethyl-N-(3-amino)propyl-ammonium bromide (PA-TELO), and N^1-(3-aminopropyl)((2-dodecyloxy)ethyl)-N^2-(2-dodecyloxy)ethyl-1-piperazinaminium bromide (GA-LOE-BP) can also be employed in the present invention.

Non-diether cationic lipids, such as DL-1,2-dioleoyl-3-dimethylaminopropyl-β-hydroxyethylammonium (DORI diester), 1-O-oleyl-2-oleoyl-3-dimethylaminopropyl-β-hydroxyethylammonium (DORI ester/ether), and their salts promote *in vivo* gene delivery. In some embodiments, cationic lipids comprise groups attached via a heteroatom attached to the quaternary ammonium moiety in the head group. A glycyl spacer can connect the linker to the hydroxyl group.

Specific, but non-limiting cationic lipids for use in certain embodiments of the present invention include DMRIE ((±)-N-(2-hydroxyethyl)-N,N-dimethyl-2,3-bis(tetradecyloxy)-1-propanaminium bromide), GAP-DMORIE ((±)-N-(3-aminopropyl)-N,N-dimethyl-2,3-bis(*syn*-9-tetradeceneyloxy)-1-propanaminium bromide), and GAP-DLRIE ((±)-N-(3-aminopropyl)-N,N-dimethyl-2,3-(*bis*-dodecyloxy)-1-propanaminium bromide).

Other cationic lipids include (±)-N,N-dimethyl-N-[2-(sperminecarboxamido)ethyl]-2,3-bis(dioleyloxy)-1-propaniminium pentahydrochloride (DOSPA), (±)-N-(2-aminoethyl)-N,N-dimethyl-2,3-bis(tetradecyloxy)-1-propaniminium bromide (β-aminoethyl-DMRIE or βAE-DMRIE) (Wheeler, *et al., Biochim. Biophys. Acta 1280*:1-11 (1996)), and (±)-N-(3-aminopropyl)-N,N-dimethyl-2,3-bis(dodecyloxy)-1-propaniminium bromide (GAP-DLRIE) (Wheeler, *et al., Proc. Natl. Acad. Sci. USA 93*:11454-11459 (1996)), which have been developed from DMRIE.

Other examples of DMRIE-derived cationic lipids that are useful for the present invention are (±)-N-(3-aminopropyl)-N,N-dimethyl-2,3-(*bis*-decyloxy)-1-propanaminium bromide (GAP-DDRIE), (±)-N-(3-aminopropyl)-N,N-dimethyl-2,3-(*bis*-tetradecyloxy)-1-propanaminium bromide (GAP-DMRIE), (±)-N-((N''-methyl)-N'-ureyl)propyl-N,N-dimethyl-2,3-bis(tetradecyloxy)-1-propanaminium bromide (GMU-DMRIE), (±)-N-(2-hydroxyethyl)-N,N-dimethyl-2,3-bis(dodecyloxy)-1-propanaminium bromide (DLRIE), and (±)-N-(2-hydroxyethyl)-N,N-dimethyl-2,3-*bis*-([Z]-9-octadecenyloxy)propyl-1- propaniminium bromide (HP-DORIE).

In the embodiments where the immunogenic composition comprises a cationic lipid, the cationic lipid may be mixed with one or more co-lipids. For purposes of definition, the term "co-lipid" refers to any hydrophobic material which may be combined with the cationic lipid component and includes amphipathic lipids, such as phospholipids, and neutral lipids, such as cholesterol. Cationic lipids and co-lipids may be mixed or combined in a number of ways to produce a variety of non-covalently bonded macroscopic

structures, including, for example, liposomes, multilamellar vesicles, unilamellar vesicles, micelles, and simple films. One non-limiting class of co-lipids are the zwitterionic phospholipids, which include the phosphatidylethanolamines and the phosphatidylcholines. Examples of phosphatidylethanolamines, include DOPE, DMPE and DPyPE. In certain embodiments, the co-lipid is DPyPE, which comprises two phytanoyl substituents incorporated into the diacylphosphatidylethanolamine skeleton.

When a composition of the present invention comprises a cationic lipid andco-lipid, the cationic lipid:co-lipid molar ratio may be from about 9:1 to about 1:9, from about 4:1 to about 1:4 , from about 2:1 to about 1:2, or about 1:1.

In order to maximize homogenity, the cationic lipid and co-lipid components may be dissolved in a solvent such as chloroform, followed by evaporation of the cationic lipid/co-lipid solution under vacuum to dryness as a film on the inner surface of a glass vessel (*e.g.*, a Rotovap round-bottomed flask). Upon suspension in an aqueous solvent, the amphipathic lipid component molecules self-assemble into homogenous lipid vesicles. These lipid vesicles may subsequently be processed to have a selected mean diameter of uniform size prior to complexing with, for example, a codon-optimized polynucleotide of the present invention, according to methods known to those skilled in the art. For example, the sonication of a lipid solution is described in Felgner *et al., Proc. Natl. Acad. Sci. USA* 84,7413-7417 (1987) and in U.S. Pat. No. 5,264,618, the disclosures of which are incorporated herein by reference.

In those embodiments where the composition includes a cationic lipid, polynucleotides of the present invention are complexed with lipids by mixing, for example, a plasmid in aqueous solution and a solution of cationic lipid:co-lipid as prepared herein are mixed. The concentration of each of the constituent solutions can be adjusted prior to mixing such that the desired final plasmid/cationic lipid:co-lipid ratio and the desired plasmid final concentration will be obtained upon mixing the two solutions. The cationic

lipid:co-lipid mixtures are suitably prepared by hydrating a thin film of the mixed lipid materials in an appropriate volume of aqueous solvent by vortex mixing at ambient temperatures for about 1 minute. The thin films are prepared by admixing chloroform solutions of the individual components to afford a desired molar solute ratio followed by aliquoting the desired volume of the solutions into a suitable container. The solvent is removed by evaporation, first with a stream of dry, inert gas (*e.g.* argon) followed by high vacuum treatment.

Other hydrophobic and amphiphilic additives, such as, for example, sterols, fatty acids, gangliosides, glycolipids, lipopeptides, liposaccharides, neobees, niosomes, prostaglandins and sphingolipids, may also be included in compositions of the present invention. In such compositions, these additives may be included in an amount between about 0.1 mol % and about 99.9 mol % (relative to total lipid), about 1-50 mol %, or about 2-25 mol %.

Additional embodiments of the present invention are drawn to compositions comprising an auxiliary agent. The present invention is further drawn to methods to use such compositions, methods to make such compositions, and pharmaceutical kits. As used herein, an "auxiliary agent" is a substance included in a composition for its ability to enhance, relative to a composition which is identical *except* for the inclusion of the auxiliary agent, the entry of polynucleotides into vertebrate cells *in vivo*, and/or the *in vivo* expression of polypeptides encoded by such polynucleotides. Auxiliary agents of the present invention include nonionic, anionic, cationic, or zwitterionic surfactants or detergents, in particular, nonionic surfactants or detergents, chelators, DNase inhibitors, agents that aggregate or condense nucleic acids, emulsifying or solubilizing agents, wetting agents, gel-forming agents, and buffers.

Auxiliary agents for use in compositions of the present invention include, but are not limited to non-ionic detergents and surfactants IGEPAL CA 630® CA 630, NONIDET NP-40, Nonidet ® P40, Tween-20®, Tween-80®, Pluronic® F68, Pluronic F77®, Pluronic P65®, Triton X-100™, and

Triton X-114™; the anionic detergent sodium dodecyl sulfate (SDS); the sugar stachyose; the condensing agent DMSO; and the chelator/DNAse inhibitor EDTA. In certain specific embodiments, the auxiliary agent is DMSO, Nonidet P40, Pluronic F68®, Pluronic F77®, Pluronic P65®, Pluronic L64®, and Pluronic F108®. *See, e.g.,* U.S. Patent Application Publication 20020019358, published February 14, 2002, which is incorporated herein by reference in its entirety.

Compositions of the present invention can be formulated according to known methods. Suitable preparation methods are described, for example, in Remington's Pharmaceutical Sciences, 16th Edition, A. Osol, ed., Mack Publishing Co., Easton, PA (1980), and Remington's Pharmaceutical Sciences, 19th Edition, A.R. Gennaro, ed., Mack Publishing Co., Easton, PA (1995), both of which are incorporated herein by reference in their entireties. Although the composition may be administered as an aqueous solution, it can also be formulated as an emulsion, gel, solution, suspension, lyophilized form, or any other form known in the art. In addition, the composition may contain pharmaceutically acceptable additives including, for example, diluents, binders, stabilizers, and preservatives.

Certain compositions of the present invention may further include one or more known adjuvants. The term "adjuvant" refers to any material having the ability to (1) alter or increase the immune response to a particular antigen or (2) increase or aid an effect of a pharmacological agent. It should be noted, with respect to polynucleotide vaccines, that an "adjuvant," may be a transfection facilitating material. Similarly, certain "transfection facilitating materials" described *supra*, may also be an "adjuvant." An adjuvant may be used with a composition comprising a polynucleotide of the present invention. In a prime-boost regiment, as described herein, an adjuvant may be used with either the priming immunization, the booster immunization, or both. Suitable adjuvants include, but are not limited to, cytokines and growth factors; bacterial components (*e.g.*, endotoxins, in particular superantigens, exotoxins

and cell wall components); aluminum-based salts; calcium-based salts; silica; polynucleotides; toxoids; serum proteins, viruses and virally-derived materials, poisons, venoms, and cationic lipids.

The ability of an adjuvant to increase the immune response to an antigen is typically manifested by a significant increase in immune-mediated protection. For example, an increase in humoral immunity is typically manifested by a significant increase in the titer of antibodies raised to the antigen, and an increase in T-cell activity is typically manifested in increased cell proliferation, or cellular cytotoxicity. An adjuvant may also alter an immune response, for example, by changing a primarily humoral or Th_2 response into a primarily cellular, or Th_1 response.

In certain adjuvant compostions, the adjuvants are cytokines. A composition of the present invention can comprise one or more cytokines, chemokines, or compounds that induce the production of cytokines and chemokines, or a polynucleotide encoding one or more cytokines, chemokines, or compounds that induce the production of cytokines and chemokines. Examples include, but are not limited to granulocyte macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), colony stimulating factor (CSF), erythropoietin (EPO), interleukin 2 (IL-2), interleukin-3 (IL-3), interleukin 4 (IL-4), interleukin 5 (IL-5), interleukin 6 (IL-6), interleukin 7 (IL-7), interleukin 8 (IL-8), interleukin 10 (IL-10), interleukin 12 (IL-12), interleukin 15 (IL-15), interleukin 18 (IL-18), interferon alpha (IFNα), interferon beta (IFNβ), interferon gamma (IFNγ), interferon omega (IFNω), interferon tau (IFNτ), interferon gamma inducing factor I (IGIF), transforming growth factor beta (TGF-β), RANTES (regulated upon activation, normal T-cell expressed and presumably secreted), macrophage inflammatory proteins (*e.g.*, MIP-1 alpha and MIP-1 beta), *Leishmania* elongation initiating factor (LEIF), and Flt-3 ligand.

In certain compositions of the present invention, the polynucleotide construct may be complexed with an adjuvant composition comprising (±)-N-

(3-aminopropyl)-N,N-dimethyl-2,3-bis(*syn*-9-tetradeceneyloxy)-1-propanaminium bromide (GAP-DMORIE). The composition may also comprise one or more co-lipids, *e.g.*, 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine (DOPE), 1,2-diphytanoyl-*sn*-glycero-3-phosphoethanolamine (DPyPE), and/or 1,2-dimyristoyl-glycer-3-phosphoethanolamine (DMPE). An adjuvant composition comprising ;GAP-DMORIE and DPyPE at a 1:1 molar ratio is referred to herein as Vaxfectin™. *See, e.g.*, PCT Publication No. WO 00/57917, which is incorporated herein by reference in its entirety.

Nucleic acid molecules and/or polynucleotides of the present invention, *e.g.*, pDNA, mRNA, linear DNA or oligonucleotides, may be solubilized in any of various buffers. Suitable buffers include, for example, phosphate buffered saline (PBS), normal saline, Tris buffer, and sodium phosphate (*e.g.*, 150 mM sodium phosphate). Insoluble polynucleotides may be solubilized in a weak acid or weak base, and then diluted to the desired volume with a buffer. The pH of the buffer may be adjusted as appropriate. In addition, a pharmaceutically acceptable additive can be used to provide an appropriate osmolarity. Such additives are within the purview of one skilled in the art. For aqueous compositions used *in vivo*, sterile pyrogen-free water can be used. Such formulations will contain an effective amount of a polynucleotide together with a suitable amount of an aqueous solution in order to prepare pharmaceutically acceptable compositions suitable for administration to a vertebrate.

EXAMPLES

Materials and Methods

The following materials and methods apply generally to all the examples disclosed herein. Specific materials and methods are disclosed in each example, as necessary.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology (including PCR), vaccinology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. *See*, for example, *Molecular Cloning A Laboratory Manual*, 2nd Ed., Sambrook *et al.*, ed., Cold Spring Harbor Laboratory Press: (1989); *DNA Cloning*, Volumes I and II (D. N. Glover ed., 1985); *Oligonucleotide Synthesis* (M. J. Gait ed., 1984); Mullis *et al.* U.S. Pat. No: 4,683,195; *Nucleic Acid Hybridization* (B. D. Hames & S. J. Higgins eds. 1984); *Transcription And Translation* (B. D. Hames & S. J. Higgins eds. 1984); *Culture Of Animal Cells* (R. I. Freshney, Alan R. Liss, Inc., 1987); *Immobilized Cells And Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide To Molecular Cloning* (1984); the treatise, *Methods In Enzymology* (Academic Press, Inc., N.Y.); *Gene Transfer Vectors For Mammalian Cells* (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); *Methods In Enzymology*, Vols. 154 and 155 (Wu *et al.* eds.), *Immunochemical Methods In Cell And Molecular Biology* (Mayer and Walker, eds., Academic Press, London, 1987); and in Ausubel *et al.*, *Current Protocols in Molecular Biology,* John Wiley and Sons, Baltimore, Maryland (1989).

Plasmid Vector

Constructs of the present invention were inserted into eukaryotic expression vector V1012. This vector is built on a modified pUC18 background (*see* Yanisch-Perron, C., *et al. Gene 33*:103-119 (1985)), and contains a kanamycin resistance gene, the human cytomegalovirus immediate early 1 promoter/enhancer and intron A, and the bovine growth hormone transcription termination signal, and a polylinker for inserting foreign genes. *See* Hartikka, J., *et al., Hum. Gene Ther. 7*:1205-1217 (1996). However, other standard commercially available eukaryotic expression vectors may be used in the present invention, including, but not limited to: plasmids pcDNA3,

pCMV/Zeo, pCR3.1, pEF1/His, pIND/GS, pRc/CMV2, pSV40/Zeo2, pTRACER-CMV, pUB6/V5-His, pVAX1, and pZeoSV2 (available from Invitrogen, San Diego, CA), and plasmid pCI (available from Promega, Madison, WI).

Plasmid DNA purification

Plasmid DNA was transformed into *Escherichia coli* DH5α competent cells and highly purified covalently closed circular plasmid DNA was isolated by a modified lysis procedure (Horn, N.A., *et al., Hum. Gene Ther. 6*:565-573 (1995)) followed by standard double CsCl-ethidium bromide gradient ultracentrifugation (Sambrook, J., *et al., Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, Plainview, New York (1989)). Alternatively, plasmid DNAs are purified using Giga columns from Qiagen (Valencia, CA) according to the kit instructions. All plasmid preparations were free of detectable chromosomal DNA, RNA and protein impurities based on gel analysis and the bicinchoninic protein assay (Pierce Chem. Co., Rockford IL). Endotoxin levels were measured using *Limulus* Amebocyte Lysate assay (LAL, Associates of Cape Cod, Falmouth, MA) and were less than 0.6 Endotoxin Units/mg of plasmid DNA. The spectrophotometric A_{260}/A_{280} ratios of the DNA solutions were typically above 1.8. Plasmids were ethanol precipitated and resuspended in an appropriate solution, *e.g.*, 150 mM sodium phosphate (for other appropriate excipients and auxiliary agents, see U.S. Patent Application Publication 20020019358, published February 14, 2002). DNA was stored at -20°C until use. DNA was diluted by mixing it with 300 mM salt solutions and by adding appropriate amount of USP water to obtain 1 mg/ml plasmid DNA in the desired salt at the desired molar concentration.

Injections of plasmid DNA

The quadriceps muscles of restrained awake mice (*e.g.*, female 6 - 12 week old BALB/c mice from Harlan Sprague Dawley, Indianapolis, IN) are injected bilaterally with 50 μg of DNA in 50 μl solution (100 μg in 100 μl total per mouse) using a disposable sterile, plastic insulin syringe and 28G 1/2 needle (Becton-Dickinson, Franklin Lakes, NJ, Cat. No. 329430) fitted with a plastic collar cut from a micropipette tip, all as previously described (Hartikka, J., *et al.*, *Hum. Gene Ther.* *7*:1205-1217 (1996)).

Animal care throughout the study was in compliance with the "Guide for the Use and Care of Laboratory Animals", Institute of Laboratory Animal Resources, Commission on Life Sciences, National Research Council, National Academy Press, Washington, D.C., 1996 as well as with Vical's Institutional Animal Care and Use Committee.

Immune Correlates

Since anthrax challenge experiments must be carried under strict containment conditions, they can be difficult and expensive, even in laboratory animals. Accordingly, it has been very important for workers in this area to develop *in vitro* assays to measure levels of immunity and to demonstrate that these assays sufficiently correlate to *in vivo* challenges. A number of *in vitro* assays, which are known to those of ordinary skill in the art to be correlates for challenges have been have been developed. *See, e.g.*, Reuveny, S. *et al. Infect. Immun.* *69*:2888-2893 (2001); Kobiler, D. *et al. Infect. Immun.* *70*:544-560 (2002); Pitt, M.L. *et al. Vaccine* *19*:4768-4773 (2001); and Park, S., and Leppla, S.H. *Protein Expr. Purif.* *18*:293-302 (2000), each of which is incorporated herein by reference in its entirety. An additional assay is described in Example 9(b), *infra*.

EXAMPLE 1

Construction of an Isolated Polynucleotide Comprising a Human Codon-Optimized PA Coding Region, Encoding the Full Length *Bacillus Anthracis* Protective Antigen (PA)

A representative native *Bacillus anthracis* protective antigen (PA) nucleotide sequence consists of nucleotides 1804 to 4098 of GenBank accession number M22589 version M22589.1 GI:143280 (SEQ ID NO:3). *See* Welkos,S.L. *et al. Gene 69*:287-300 (1988), which is incorporated herein by reference in its entirety. The PA sequence encodes a 764 amino acid (aa) precursor protein (SEQ ID NO:4) that is processed by a signal peptidase upon secretion by the bacteria, and also by host serum proteases (reviewed in Mesnage S., and Fouet, A. *J. Bacteriol. 184*:331-334 (2002), which is incorporated by reference herein in its entirety). The first 29 amino acids of PA encodes a bacterial signal sequence that is cleaved during secretion from the bacteria. In the host, furin-like serum proteases cleave off the N-terminal 258 amino acids to yield PA63, the active form of PA that can bind lethal factor (LF) and edema factor (EF), thereby causing toxicity.

A nucleic acid coding region for full-length PA (SEQ ID NO:4), optimized for human codon usage was derived by determining codon frequencies from the human codon usage table (Table 2) as described above. The codon-optimized nucleic acid sequence was created by using the various codons encoding the amino acids of SEQ ID NO:4, each at the frequencies with which they occur in the codon usage table of Table 2. Although any codon-optimized coding region which encodes SEQ ID NO:4 may be used, including, but not limited to SEQ ID Nos 23, 24, or 25, this Example and other Examples below use the human codon-optimized coding region encoding SEQ ID NO:4 represented by SEQ ID NO:23. Alternatively a human codon-optimized nucleic acid coding region encoding SEQ ID NO:4 can be prepared by referring to the codon usage table of Table 2, and using only the most frequent codons for each amino acid, as represented by SEQ ID NO:21.

The nucleic acid represented by SEQ ID NO:23 is constructed in the following manner. First, a series complementary oligonucleotide pairs of 80-90 nucleotides each in length and spanning the length of SEQ ID NO:23 are synthesized by standard methods. These oligonucleotide pairs are synthesized such that upon annealing, they form double stranded fragments of 80-90 base pairs, containing cohesive ends. The single-stranded ends of each pair of oligonucleotides are designed to anneal with a single-stranded end of an adjacent oligonucleotide duplex. Several adjacent oligonucleotide pairs prepared in this manner are allowed to anneal, and approximately five to six adjacent oligonucleotide duplex fragments are then allowed to anneal together via the cohesive single stranded ends. This series of annealed oligonucleotide duplex fragments is then ligated together and cloned into the TOPO® vector available from Invitrogen Corporation, Carlsbad, CA. The construct is then sequenced by standard methods. Constructs prepared in this manner, comprising 5 to 6 adjacent 80 to 90 base pair fragments ligated together, *i.e.*, fragments of about 500 base pairs, are prepared, such that the entire desired sequence of SEQ ID NO:23 is represented in a series of plasmid constructs. The inserts of these plasmids are then cut with appropriate restriction enzymes and ligated together to form the final construct. The final construct is then cloned into a standard bacterial cloning vector, and sequenced.

EXAMPLE 2

Construction of an Isolated Polynucleotide Comprising a Human Codon-Optimized LF Coding Region, Encoding the Full Length *Bacillus Anthracis* Lethal Factor (LF)

A representative native *Bacillus anthracis* lethal factor (LF) nucleotide sequence consists of nucleotides 685 to 3111 of GenBank accession number M30210 version M30210.1 GI:143141 (SEQ ID NO:11). The LF sequence encodes a 809 amino acid precursor protein that is processed to a 775 amino acid secreted protein by cleavage of its signal sequence. LF is a zinc

metalloprotease that cleaves mitogen-activated protein kinase kinases (MAPKKs) contained inside target cells. *See* Mesnage S., and Fouet, A. *J. Bacteriol. 184*:331-334 (2002). Numerous mutations in LF have been described that eliminate zinc binding or the catalytic site of LF resulting in the loss of toxicity. *See* Hammond, S.E., and Hanna, P.C. *Infect. Immun. 66*:2374-2378 (1998). One form of inactive LF is described in detail herein, but all others could also be used with an identical approach.

A nucleic acid coding region for full-length LF (SEQ ID NO:12), optimized for human codon usage was derived by determining codon frequencies from the human codon usage table (Table 2) as described above. The codon-optimized nucleic acid sequence was created by using the various codons encoding the amino acids of SEQ ID NO:12, each at the frequencies with which they occur in the codon usage table of Table 2. Although any codon-optimized coding region which encodes SEQ ID NO:12 may be used, including, but not limited to SEQ ID NOs 26, 27, and 28, this Example and other Examples below use the human codon-optimized coding region encoding SEQ ID NO:12 represented by SEQ ID NO:26. Alternatively a human codon-optimized nucleic acid coding region encoding SEQ ID NO:12 can be prepared by referring to the codon usage table of Table 2, and using only the most frequent codons for each amino acid, as represented by SEQ ID NO:22.

The nucleic acid represented by SEQ ID NO:26 is constructed commercially by Retrogen, San Diego, CA, in the following manner. First, a series complementary oligonucleotide pairs of 80-90 nucleotides each in length and spanning the length of SEQ ID NO:26 are synthesized by standard methods. These oligonucleotide pairs are synthesized such that upon annealing, they form double stranded fragments of 80-90 base pairs, containing cohesive ends. The single-stranded ends of each pair of oligonucleotides are designed to anneal with a single-stranded end of an adjacent oligonucleotide duplex. Several adjacent oligonucleotide pairs prepared in this manner are allowed to anneal, and approximately five to six

adjacent oligonucleotide duplex fragments are then allowed to anneal together via the cohesive single stranded ends. This series of annealed oligonucleotide duplex fragments are then ligated together and cloned into a the TOPO® vector available from Invitrogen Corporation, Carlsbad, CA. The construct is then sequenced by standard methods. Constructs prepared in this manner, comprising 5 to 6 adjacent 80 to 90 base pair fragments ligated together, *i.e.*, fragments of about 500 base pairs, are prepared, such that the entire desired sequence of SEQ ID NO:26 is represented in a series of plasmid constructs. The inserts of these plasmids are then cut with appropriate restriction enzymes and ligated together to form the final construct. The final construct is then cloned into a standard bacterial cloning vector, and sequenced.

EXAMPLE 3

Construction of Plasmid Constructs Comprising Fragments, Variants, and Derivatives of a Human Codon-Optimized Coding Region Encoding *Bacillus Anthracis* PA

Several fragments, variants, and derivatives based on SEQ ID NO:23, the human codon-optimized coding region encoding *Bacillus anthracis* PA described in Example 1, were constructed in the following manner. Codon-optimized nucleic acid fragments encoding three alternate forms of PA were constructed, namely, a nucleic acid fragment encoding full-length PA minus the furin cleavage site (PA83Δ Furin), a nucleic acid fragment encoding the active furin cleavage product of mature PA (PA63), and a nucleic acid fragment encoding the active furin cleavage product of mature PA in which Phe 342 and 343 have been deleted (PA63ΔFF). Each of these nucleic acid fragments were fused in-frame to a nucleic acid encoding a human tissue plasminogen activator (TPA) signal peptide sequence that directs the expressed PA variants and/or fragments to the secretory pathway in mammalian cells. Other useful PA fragments, variants and/or derivatives will

be readily apparent to those of ordinary skill in the art, and are included in the present invention.

a) Construction of TPA-PA63.

PA63, the C-terminal fragment of PA corresponding to amino acids 199-764 of SEQ ID NO:4 corresponds to the mature, extracellularly processed protein that is able to bind to LF and edema factor (EF).

TPA-PA63 (Fig. 1, SEQ ID NO:1) was constructed commercially by Retrogen, San Diego, CA. A large number of other companies which provide similar construction of predetermined nucleic acid sequences are well known to those of ordinary skill in the art. The sequence was constructed in the following manner. First, a series complementary oligonucleotide pairs of 80-90 nucleotides each in length and spanning the length of SEQ ID NO:1 were synthesized by standard methods. These oligonucleotide pairs were synthesized such that upon annealing, they formed double stranded fragments of 80-90 base pairs, containing cohesive ends. The single-stranded ends of each pair of oligonucleotides were designed to anneal with a single-stranded end of an adjacent oligonucleotide duplex. Several adjacent oligonucleotide pairs prepared in this manner were allowed to anneal, and approximately five to six adjacent oligonucleotide duplex fragments were then allowed to anneal together via the cohesive single stranded ends. This series of annealed oligonucleotide duplex fragments were then ligated together and cloned into a the TOPO® vector available from Invitrogen Corporation, Carlsbad, CA. The construct was then sequenced by standard methods. Constructs prepared in this manner, comprising 5 to 6 adjacent 80 to 90 base pair fragments ligated together, *i.e.*, fragments of about 500 base pairs, were prepared, such that the entire desired sequence of SEQ ID NO:1 was represented in a series of plasmid constructs. The inserts of these plasmids were then cut with appropriate restriction enzymes and ligated together in the TOPO® vector.

This construct was cut with EcoRV + BamHI and the 1788 bp insert fragment (*i.e.*, SEQ ID NO:1) was cloned into the same sites of the VR1012 expression plasmid (see Hartikka *et al.*, *Hum. Gene Therapy* *7*:1205-1217 (1996), which is incorporated herein by reference in its entirety). The resulting plasmid, designated VR6290, was sequenced and expressed in transiently transfected VM-92 cells in culture (see Example 6) to confirm the expression and secretion of the construct.

b) Construction of TPA-PA63ΔFF.

A different non-toxic form of PA can be generated by deleting the two phenylalanine residues at positions 342 and 343 of SEQ ID NO:4 to produce a PA protein that cannot heptamerize and form a pore to allow LF and EF to enter the cytoplasm of an infected cell. *See, e.g.*, Singh, Y. *et al. J. Biol. Chem. 269*:29039-29046 (1994), which is incorporated herein by reference in its entirety.

An expression plasmid comprising TPA-PA63Δ FF (Fig. 2, SEQ ID NO:5) was prepared by the following method. Plasmid VR6290, prepared as described in section (a), *supra*, was used as a template for PCR with the following two sets of PCR primers using Turbo *Pfu* polymerase from Stratagene Inc., La Jolla, CA

1. TPA –for 5'GAGCTTGATA TCGCCACCAT GGATGC 3' (SEQ ID NO:29) and PA del FF-Rev 5' CCACCAATAT CCGATGCATG GACTTCCGC 3' (SEQ ID NO:30) produced a 520 bp fragment.

2. HPA-endRev 5' CTTGAAGGAT CCTCAACCGA TCTCGTAG 3' (SEQ ID NO:31) and PA del FF-For 5' CCATGCATCG GATATTGGTG GCTCCGTGTC 3' (SEQ ID NO:32) produced a 1280 bp fragment that overlapped fragment 1.

Fragments 1 and 2 were gel purified using the QIAquick Gel Extraction Kit from Qiagen Inc (Valencia, CA) and the fragments were combined in a subsequent PCR reaction and amplified with the primer pair

TPA-for and HPA-endRev to yield the full length 1782 bp fragment shown in Fig 2. This fragment was digested with the restriction enzymes EcoR5 + BamHI and ligated into the same sites of the VR1012 expression plasmid. The resulting plasmid, designated VR6291, was sequenced and expressed in transiently transfected VM-92 cells in culture (see Example 6) to confirm the expression and secretion of the construct.

c) Construction of TPA-PA83Δ Furin.

Deleting the furin cleavage site of the mature PA, *i.e.*, amino acids 192-197 (Ser-Arg-Lys-Lys-Arg-Ser) of SEQ ID NO:4, yields a protein that is secreted from the cell and that can bind the host cell receptor but cannot bind LF or EF and therefore is non-toxic. *See*, *e.g.*, Singh, Y. *et al. Infect. Immun.* *66*:3447-3448 (1998), and Klimpel KR, *et al. Proc. Natl. Acad. Sci. USA* *89*:10277-10281 (1992), which are incorporated herein by reference in their entireties.

An expression plasmid comprising TPA-PA83Δfurin (Fig. 3, SEQ ID NO:7) was constructed in the following manner. A plasmid comprising a codon-optimized nucleotide sequence (as per the sequence of SEQ ID NO:23) encoding the N-terminal 20 kD domain of PA, *i.e.*, corresponding to the portion of PA that is cleaved off by furin, was synthesized by Retrogen Inc. according to the method described in section a), *supra*. This plasmid was cut with EcoRV+AfeI and the 570 bp insert was gel purified as above. The plasmid VR6290 described in section a) above was digested with EcoRV+AfeI and the 6.6 kb linear vector fragment was gel purified and ligated to the 570 bp N-terminal fragment. Transformed colonies were screened for recombinants by PCR using the primer pair NtermPA seqF 5' GTGGACGACC AGGAAGTGAT C 3' (SEQ ID NO:33) and NtermPA seqR 5' GGCTATCTGT CCAGTACAGC TTGAA3' (SEQ ID NO:34). A selected recombinant, designated VR6292, was sequenced and was expressed in

transiently transfected VM-92 cells in culture (see Example 6) to confirm the expression and secretion of the construct.

EXAMPLE 4

Construction of Plasmid Constructs Comprising Fragments, Variants, and Derivatives of a Human Codon-Optimized Coding Region Encoding *Bacillus Anthracis* LF

Several fragments, variants, and derivatives of SEQ ID NO:26, the human codon-optimized coding region encoding *Bacillus anthracis* LF, as prepared in Example 2, were constructed in the following manner. Codon-optimized nucleic acid fragments encoding four alternate forms of LF were constructed, namely, a nucleic acid fragment encoding the full-length mature LF in which His 686, His 690 and Glu 687 have been substituted with Ala, Ala, and Asp, respectively (LF HEXXH), a nucleic acid fragment encoding amino acids 34 to 583 of full-length LF, encoding domains I-III of mature LF (LF Domain I-III), a nucleic acid fragment encoding amino acids 34 to 254 of mature LF, corresponding to domain I of mature LF (LF Domain IA), and a nucleic acid fragment encoding amino acids 34 to 295 of mature LF, corresponding to domain I of mature LF (LF Domain IB). Each of these nucleic acid fragments were fused in-frame to a nucleic acid encoding a human tissue plasminogen activator (TPA) signal peptide sequence that directs the expressed LF variants and/or fragments to the secretory pathway in mammalian cells. Furthermore, other useful LF fragments, variants and/or derivatives would be readily apparent to those of ordinary skill in the art.

a) Construction of TPA-LF HEXXH.

This construct encodes full length LF (minus the bacterial signal sequence) with three point mutations that render LF non-toxic. Each of these mutations, alone or together, are thought to eliminate the enzymatic activity of LF, thereby rendering it non-toxic. See, *e.g.*, Hammond, SE and Hanna PC,

Infect. Immun. *66*:2374-2378 (1998), which is incorporated herein by reference in its entirety. Other LF mutants contained in this reference, *e.g.*, LF^{E687C}, LF^{E687D}, LF^{H686A}, LF^{H690A}, and $LF^{H686A+H690A}$, are also included in the present invention. While not being bound by theory, substitution of the histidine residues at positions 686 and 690 is thought to decrease zinc binding, resulting in decreased or no protease activity, and substitution of the glutamic acid at position 687 is thought to also eliminate protease activity, thereby resulting in no *in vitro* or *in vivo* macrophage killing. This construct combines all three mutations to afford a greater perceived level of safety than either point mutation alone.

An expression plasmid comprising LF HEXXH (Fig. 4, SEQ ID NO:9) was prepared in the following manner. The entire 2418 bp sequence was synthesized by Retrogen Inc. and inserted into the EcoRV and BamHI sites of the TOPO vector as described in Example 3(a). The resulting plasmid was digested with EcoRV and BamHI and the 2418 bp insert was purified by gel electrophoresis as described above. The insert was ligated into EcoRV+BamHI digested VR1012 and transformed into *E. coli*. Transformed colonies were screened for recombinants by PCR using the primer pair seqF1-hLF 5' CCGTGCTCGT TATTCAGAGT 3' (SEQ ID NO:35) and seqR2-hLF 5' CCTTCTCTTC TGTGCTAAGG 3' (SEQ ID NO:36). A selected recombinant, designated VR6295, was sequenced and was expressed in transiently transfected VM-92 cells in culture (see Example 6) to confirm the expression and secretion of the construct.

b) Construction of TPA-LF Domain I-III.

This construct encodes the N-terminal amino acids 34-583 of mature LF, corresponding to domains I-III. The entire protease domain (domain IV) has been deleted and is therefore non-toxic. *See, e.g.*, Pannifer AD *et al.* *Nature* *414*:229-233 (2001), which is incorporated herein by reference in its entirety.

Recent data suggest LF is capable of entering cells independently of PA using a region at the N terminal domain (Kushner, N. *et al. Proc. Natl. Acad. Sci. 100*: 6652-7 (2003)). In addition, the full length LF is able to cause impairment of dendritic cell function via domain IV protease degredation of MAP activated protein kinase kinase (MAPKK). Agrawal, A. *et al. Nature 424*: 329-34 (2003)). Therefore LF, independently of its association with PA, may have toxic effects which could be blocked through vaccination. Through the inclusion of an LF component in the vaccine of the present invention, it may be possible to neutralize LF at a number of domains and to block potential toxicities that occur in conjunction with, or independent of, binding to PA. In may also be possible to block the primary binding of LF to PA.

An expression plasmid comprising TPA-LF Domain I-III (Fig. 5, SEQ ID NO:13) was prepared in the following manner. The plasmid VR6295 (as produced in section a) above, was PCR amplified with the primer pair TPA–for (SEQ ID NO:29) and LF-DomII-R 5' GAACCTGGAT CCCTACACCA CCTTGGCGTC GATG 3' (SEQ ID NO:37) using *Pfu* polymerase. The 1740 bp fragment was gel purified, digested with EcoRV + BamHI and cloned into VR1012. Transformed colonies were screened by PCR using the same amplification primers. A selected recombinant, designated VR62952, was sequenced and was expressed in transiently transfected VM-92 cells in culture (see Example 6) to confirm the expression and secretion of the construct.

c) Construction of TPA-LF Domain IA.

This construct encodes the N-terminal amino acids 34-254 of mature LF, corresponding generally to domain I. This is the portion of LF that directly binds PA. *See, e.g.*, Pannifer AD *et al. Nature 414*:229-233 (2001).

An expression plasmid comprising TPA-LF Domain I (Fig. 6, SEQ ID NO:15) was prepared in the following manner. The plasmid VR6295 (as produced in section a) above, was PCR amplified with the primer pair TPA–for (SEQ ID NO:29) and G-LF-R 5'GCTAATGGAT CCTCAAAATG

CCTTGGCGAA CACCT 3' (SEQ ID NO:38) using *Pfu* polymerase. The 753 bp fragment was gel purified, digested with EcoRV + BamHI and cloned into VR1012. Transformed colonies were screened by PCR using the same amplification primers. A selected recombinant, designated VR6295G, was sequenced and was expressed in transiently transfected VM-92 cells in culture (see Example 6) to confirm the expression and secretion of the construct.

d) Construction of TPA-LF Domain IB.

This construct encodes the N-terminal amino acids 34-295 of mature LF, also corresponding generally to domain I.

An expression plasmid comprising TPA-LF Domain IB (Fig. 14, SEQ ID NO:39) was prepared in the following manner. The plasmid VR6295 (as produced in section a) above, was PCR amplified with the primer pair TPA–for (SEQ ID NO:29) and crystal-LF-R 5' CCATACGGAT CCTCACTGGT CTTTCAGTTC CTCCA 3' (SEQ ID NO:41) using *Pfu* polymerase. The 876 bp fragment was gel purified, digested with EcoRV + BamHI and cloned into VR1012. Transformed colonies were screened by PCR using the same amplification primers. A selected recombinant, designated VR62951, was sequenced and was expressed in transiently transfected VM-92 cells in culture (see Example 6) to confirm the expression and secretion of the construct.

EXAMPLE 5

N-Linked Glycosylation Mutants

Most mammalian transmembrane and secreted proteins are glycosylated post-translationally in the endoplasmic reticulum. *See, e.g.*, Lodish H *et al.* Molecular Cell Biology 4th edition, W. H. Freeman and Company, New York. There are two main types of protein glycosylation in mammalian cells, N-linked and O-linked. N-linked glycosylation occurs on asparigine (N) residues at the amino acid motif N-X-(S/T) where X refers to

any amino acid residue and S/T denotes serine or threonine. There are seven N-linked glycosylation motifs in mature LF, twelve N-linked glycosylation motifs in full-length mature PA (PA83) and ten N-linked glycosylation motifs in PA63. Since this glycosylation does not occur in bacteria, the anthrax antigens synthesized in mammalian cells after DNA immunization may differ from the PA and LF in anthrax toxin secreted by *B. anthracis*. *See, e.g.*, Schaffer C. *et al. Proteomics 1*:248-246 (2001), which is incorporated herein by reference in its entirety. This mammalian N-linked glycosylation could obscure or alter B-cell antibody epitopes that are normally exposed in conventional anthrax protein vaccines. Therefore codon-optimized coding regions encoding PA63 and LF were made in which the asparagines in the N-linked glycosylation motifs (N-X-S/T) motifs were mutated to glutamines (Q-X-S/T). The motif Q-X-S/T is not subject to glycosylation in mammals. Such "sugar minus" variants of any of the variants, fragments, derivatives, or full length coding regions disclosed herein, as well as addition variants, fragments and derivatives known to those of skill in the art are encompassed by the present invention.

All asparagine (N) residues in the N-X-S/T motifs contained in TPA-PA63 (SEQ ID NO:1, produced as described in Example 3(a)) and TPA-LFΔHEXXH (SEQ ID NO:9, produced as described in Example 4(a)) were mutated to Glutamine (Q) by generating a series of overlapping PCR fragments. As described in more detail below, these fragments were added together two at a time and amplified with primers at the extreme end of the two fragments to build larger and larger PCR fragments until a full length mutant was obtained. In each case, the full-length fragment was gel purified, digested with EcoRV + BamHI and cloned into VR1012. All PCR reactions were performed with *Pfu* polymerase from Stratagene Inc using standard conditions.

a) Construction of TPA-sugar minus PA63.

This construct is the same as SEQ ID NO:1, except that all ten N-X-S/T motifs in the encoded polypeptide have been changed to Q-X-S/T, via point mutations. The mutated construct was assembled from overlapping PCR fragments using SEQ ID NO:1 as the template.

Ten nanograms quantities of plasmidVR6290 DNA was amplified with each of the 10 primer pairs listed in Table 8. The fourth column lists the size of the resulting PCR products with the various primer pairs. Each of these resulting PCR fragments has a single stranded region at each end, which can anneal with a single stranded region on another of the fragments.

TABLE 8

PCR Fragment	Forward Primer	Reverse Primer	Size
1	TPA-For (SEQ ID NO:29)	PA-R1 (SEQ ID NO:42)	310 bp
2	PA-F2 (SEQ ID NO:43)	PA-R2 (SEQ ID NO:44)	140 bp
3	PA-F3 (SEQ ID NO:45)	PA-R3 (SEQ ID NO:46)	90 bp
4	PA-F4 (SEQ ID NO:47)	PA-R4 (SEQ ID NO:48)	180 bp
5	PA-F5 (SEQ ID NO:49)	PA-R5 (SEQ ID NO:50)	260 bp
6	PA-F6 (SEQ ID NO:51)	PA-R6 (SEQ ID NO:52)	100 bp
7	PA-F7 (SEQ ID NO:53)	PA-R7 (SEQ ID NO:54)	185 bp
8	PA-F8 (SEQ ID NO:55)	PA-R8 (SEQ ID NO:56)	150 bp
9	PA-F9 (SEQ ID NO:57)	PA-R9 (SEQ ID NO:58)	130 bp
10	PA-F10 (SEQ ID NO:59)	PA-R10 (SEQ ID NO:60)	135 bp
11	PA-F11 (SEQ ID NO:61)	HPA-endRev (SEQ ID NO:31)	80 bp

2.5 microliters of each PCR fragment in Table 8 was combined pair wise with a second PCR fragment in Table 8. The two fragments were allowed to anneal and were used as templates in a second series of PCR reactions, with resulting PCR fragments as shown in Table 9.

TABLE 9

PCR Fragment	Template Fragments	Forward Primer	Reverse Primer	Size
12	2 + 3	PA-F2 (SEQ ID NO:43)	PA-R3 (SEQ ID NO:46)	230 bp
13	4 + 5	PA-F4 (SEQ ID NO:47)	PA-R5 (SEQ ID NO:50)	440 bp
14	6 + 7	PA-F6 (SEQ ID NO:51)	PA-R7 (SEQ ID NO:54)	285 bp
15	8 + 9	PA-F8 (SEQ ID NO:55)	PA-R9 (SEQ ID NO:58)	280 bp
16	10 + 11	PA-F10 (SEQ ID NO:59)	HPA-endRev (SEQ ID NO:31)	215 bp

2.5 microliters of each PCR fragment in Table 9 was combined pair wise with a second PCR fragment in Table 9. The two fragments were allowed to anneal and were used as templates in a third series of PCR reactions, with resulting PCR fragments as shown in Table 10.

TABLE 10

PCR Fragment	Template Fragments	Forward Primer	Reverse Primer	Size
17	1 + 12	TPA-For (SEQ ID NO:29)	PA-R3 (SEQ ID NO:46)	540 bp
18	13 + 14	PA-F4 (SEQ ID NO:47)	PA-R7 (SEQ ID NO:54)	725 bp
19	15 + 16	PA-F8 (SEQ ID NO:55)	HPA-endRev (SEQ ID NO:31)	495 bp

Fragments 17, 18, and 19 were gel purified before proceeding to the next series of PCR reactions. The last two sets of PCR reactions were carried out as listed in Table 11, using 2.5 microliters of the annealed PCR fragment pairs listed in the second column, which had been gel purified.

TABLE 11

PCR Fragment	Template Fragments	Forward Primer	Reverse Primer	Size
20	17 + 18	TPA-For (SEQ ID NO:29)	PA-R7 (SEQ ID NO:54)	1265 bp
21	19 + 20	TPA-For (SEQ ID NO:29)	HPA-endRev (SEQ ID NO:31)	1788 bp

Resulting PCR fragment 21 represents the full-length TPA-Sugar minus PA63 fragment (Figure 7, SEQ ID NO:17). The TPA-sugar minus PA63 fragment was cloned into the VR1012 expression plasmid. A selected recombinant, designated VR6299, was sequenced, and was expressed in

transiently transfected VM-92 cells in culture (see Example 6) to confirm the expression and secretion of the construct.

The sequences of the primers used in the PCR reactions in this Example are listed in Table 12.

TABLE 12

SEQ ID NO:	Primer	Sequence
29	TPA –for	GAGCTTGATATCGCCACCATGGATGC
42	PA-R1	CTGGAGACACCTGTTTATCGATCC
43	PA-F2	GGATCGATAAACAGGTGTCTCCAG
44	PA-R2	GAAGTACTGGTCTGTTTAGATATGGT
45	PA-F3	ACCATATCTAAACAGACCAGTACTTC
46	PA-R3	CGTCGAGGACTGGCTATTGCTAA
47	PA-F4	TTAGCAATAGCCAGTCCTCGACG
48	PA-R4	GAGGGTCTGCTGTTTGCCCAGG
49	PA-F5	CCTGGGCAAACAGCAGACCCTC
50	PA-R5	CTTCAGACCACTGTGACCCAGTG
51	PA-F6	CACTGGGTCACAGTGGTCTGAAG
52	PA-R6	GATCACTGGGCTGCACGGCGG
53	PA-F7	CCGCCGTGCAGCCCAGTGATC
54	PA-R7	TATTGGTGGCCTGCAGCTCTGC
55	PA-F8	GCAGAGCTGCAGGCCACCAATA
56	PA-R8	CAGTACTGCTCTGGATAACTTCCC
57	PA-F9	GGGAAGTTATCCAGAGCAGTACTG
58	PA-R9	AAGCTGGAAATCTGCAGCATATCAT
59	PA-F10	ATGATATGCTGCAGATTTCCAGCTT
60	PA-R10	CTCGCTTGGCTGGATGATTGTGT
61	PA-F11	ACACAATCATCCAGCCAAGCGAG
31	HPA-endRev	CTTGAAGGATCCTCAACCGATCTCGTAG

b) Construction of TPA-sugar minus LF HEXXH.

This construct is the same as SEQ ID NO:9, except that all seven N-X-S/T motifs in the encoded polypeptide have been changed to Q-X-S/T, via point mutations. The mutated construct was assembled from overlapping PCR fragments using standard methods, using primers which code for Q residues instead of N residues in the seven glycosylation motifs, and using SEQ ID NO:9 as the template.

Ten nanograms quantities of plasmidVR6295 DNA was amplified with each of the 8 primer pairs listed in Table 13. The fourth column lists the size of the resulting PCR products with the various primer pairs. Each of these resulting PCR fragments has a single stranded region at each end, which can anneal with a single stranded region on another of the fragments.

TABLE 13

PCR Fragment	Forward Primer	Reverse Primer	Size
1	TPA-For (SEQ ID NO:29)	LF-R1 (SEQ ID NO:62)	170 bp
2	LF-F2 (SEQ ID NO:63)	LF -R2 (SEQ ID NO:64)	450 bp
3	LF -F3 (SEQ ID NO:65)	LF -R3 (SEQ ID NO:66)	220 bp
4	LF -F4 (SEQ ID NO:67)	LF -R4 (SEQ ID NO:68)	580 bp
5	LF -F5 (SEQ ID NO:69)	LF -R5 (SEQ ID NO:70)	700 bp
6	LF -F6 (SEQ ID NO:71)	LF -R6 (SEQ ID NO:72)	70 bp
7	LF -F7 (SEQ ID NO:73)	LF -R7 (SEQ ID NO:74)	65 bp
8	LF -F8 (SEQ ID NO:75)	HLFend-R (SEQ ID NO:76)	165 bp

2.5 microliters of each PCR fragment in Table 13 was combined pair wise with a second PCR fragment in Table 13. The two fragments were allowed to anneal and were used as templates in a second series of PCR reactions, with resulting PCR fragments as shown in Table 14.

TABLE 14

PCR Fragment	Template Fragments	Forward Primer	Reverse Primer	Size
9	1 + 2	TPA-For (SEQ ID NO:29)	LF -R2 (SEQ ID NO:64)	620 bp
10	3 + 4	LF-F3 (SEQ ID NO:65)	LF-R4 (SEQ ID NO:68)	800 bp
11	5 + 6	LF-F5 (SEQ ID NO:69)	LF-R6 (SEQ ID NO:72)	770 bp
12	7 + 8	LF-F7 (SEQ ID NO:73)	HLFend-R (SEQ ID NO:76)	230 bp

2.5 microliters of each PCR fragment in Table 14 was combined pair wise with a second PCR fragment in Table 14. The two fragments were allowed to anneal and were used as templates in a third series of PCR reactions, with resulting PCR fragments as shown in Table 15.

TABLE 15

PCR Fragment	Template Fragments	Forward Primer	Reverse Primer	Size
13	9 + 10	TPA-For (SEQ ID NO:29)	LF-R4 (SEQ ID NO:68)	1420 bp
14	11 + 12	LF-F5 (SEQ ID NO:69)	HLFend-R (SEQ ID NO:76)	1000 bp

Fragments 13 and 14 were gel purified before proceeding to the final PCR reaction. This PCR reaction was carried out as listed in Table 16, using 2.5 microliters of the annealed PCR fragment pairs listed in the second column, which had been gel purified.

TABLE 16

PCR Fragment	Template Fragments	Forward Primer	Reverse Primer	Size
15	13 + 14	TPA-For (SEQ ID NO:29)	HLFend-R (SEQ ID NO:76)	2418 bp

Resulting PCR fragment 15 represents the full-length TPA-Sugar minus LF HEXXH fragment (Figure 8, SEQ ID NO:19). The TPA-sugar minus LF HEXXH fragment was cloned into the VR1012 expression plasmid. A selected recombinant, designated VR6300, was sequenced and was expressed in transiently transfected VM-92 cells in culture (see Example 6) to confirm the expression and secretion of the construct.

The sequences of the primers used in the PCR reactions in this Example are listed in Table 17.

TABLE 17

SEQ ID NO:	Primer	Sequence
29	TPA –for	GAGCTTGATATCGCCACCATGGATGC
62	LF-R1	TCCTGTGTTTTCTGACGTTCTTCG
63	LF-F2	CGAAGAACGTCAGAAAACACAGGA
64	LF-R2	TATCTGACGCCTGTTTGATTGTGTT
65	LF-F3	AACACAATCAAACAGGCGTCAGATA
66	LF-R3	CCAGAGACAGCTGAATCTCCTGTT
67	LF-F4	AACAGGAGATTCAGCTGTCTCTGG
68	LF-R4	AGCGGTGAGCTGGTTAATATTCATG

SEQ ID NO:	Primer	Sequence
69	LF-F5	CATGAATATTAACCAGCTCACCGCT
70	LF-R5	CCTCAGAATCCTGTCGAAGCTCA
71	LF-F6	TGAGCTTCGACAGGATTCTGAGG
72	LF-R6	GATCAGACTGCTGCTTATCCAACA
73	LF-F7	TGTTGGATAAGCAGCAGTCTGATC
74	LF-R7	AGGAAGTCAGCTGACTCCCTTCC
75	LF-F8	GGAAGGGAGTCAGCTGACTTCCT
76	HLFend-R	GCAGATCTGGATCCTCAAGAG

EXAMPLE 6

In vitro Expression of Human Codon-Optimized Coding Regions Encoding *B. Anthracis* PA and LF, and Fragments, Variants and Derivatives thereof, in a Murine Cell Line

The expression plasmids described Examples 3-5 above and the corresponding wild type *Bacillus anthracis* genes were initially analyzed in *in vitro* transiently transfected cells in culture. Initial studies were carried out in a well characterized mouse melanoma cell line (VM-92, also known as UM-449), using cationic lipid-based transfection procedures well known to those of skill in the art. Other standard cell lines, for example, COS-1 cells, COS -7 cells, CHO cells, HEK-293 cells, and HeLa cells, may be used for transient transfections as well. Following transfection, cell lysates and culture supernatants of transfected cells were evaluated to compare relative levels of expression of *B. anthracis* antigen proteins. The samples were assayed by western blots and ELISAs, using commercially available anti-PA and Anti-LF monoclonal antibodies (available from Research Diagnostics Inc., Flanders NJ), so as to compare both the quality and the quantity of expressed antigen. Additionally, *in vitro* transfection assays were used to determine the effect of mixing the various plasmids comprising codon-optimized coding regions encoding non-toxic PA and LF on levels of expression in mammalian cells.

Expression products derived from cells transfected with the various polynucleotide constructs are examined to ensure the correct or predicted

molecular weight of the recombinant antigens, and immunoreactivity of the recombinant antigens (*i.e.*, to react with *B. anthracis* antisera). In addition, a comparison of expression levels (both intra- and extra-cellular) of each class of expression plasmid (*e.g.*, wild type vs. human codon-optimized; truncated vs. full-length) is made.

EXAMPLE 7

In vitro Expression of Human Codon-Optimized Coding Regions Encoding *B. Anthracis* PA and LF, and Fragments, Variants and Derivatives Thereof, in a Human Cell Line

The expression plasmids described Examples 3-5 above and the corresponding wild type *Bacillus anthracis* genes are also analyzed in *in vitro* transfected human cells in culture. These studies are carried out in a well characterized human cell line, *e.g.*, HeLa cells, ATCC Accession No. CCL-2, available from the American Type Culture Collection, Manassas, VA, using cationic lipid-based transfection procedures well known to those of skill in the art. Following transfection, cell lysates and culture supernatants of transfected cells are evaluated to compare relative levels of expression of *B. anthracis* antigen proteins. The samples are assayed by western blots and ELISAs, using commercially available antiPA and Anti-LF monoclonal antibodies (available from Research Diagnostics Inc., Flanders NJ), so as to compare both the quality and the quantity of expressed antigen. Additionally, *in vitro* transfection assays are used to determine the effect of mixing the various plasmids comprising codon-optimized coding regions encoding non-toxic PA and LF on levels of expression in human cells.

Expression products from the derived from human cells transfected with the various polynucleotide constructs are examined for molecular weight, and expression immunoreactive antigens (*i.e.*, to react with *B. anthracis* antisera). In addition, a comparison of expression levels (both intra- and

extra-cellular) of each class of expression plasmid (*e.g.*, wild type vs. human codon-optimized; truncated vs. full-length) is made.

EXAMPLE 8

Animal Immunization and Challenge

The immunogenicity of expression products encoded by the codon-optimized polynucleotides described in Examples 1-5 are evaluated based on each plasmid's ability to mount a humoral immune response *in vivo*. Plasmids are tested individually and in combinations by injecting single constructs as well as multiple constructs in various animals as described below. Immunizations are initially carried out in mice by intramuscular (IM) injections. Serum is collected from immunized animals, and the immune response is quantitated by ELISA assay using commercially available antiPA and Anti-LF monoclonal antibodies (available from Research Diagnostics Inc., Flanders NJ) according to standard protocols. The tests of immunogenicity further include measuring antibody titer, neutralizing antibody titer, and challenging immunized animals with toxin protein.

Testing in rabbits are then used to confirm the results in mice and thereby provide efficacy data for the best plasmids in more than one mammalian immunogenicity model system. Serum is collected from immunized rabbits, and antibody titers and neutralizing antibody titers are determined. In addition, immunized rabbits are tested with a spore inhalation challenge. The combined results determine the plasmids to be subsequently tested in non-human primates.

a) Mouse immunizations.

The plasmid constructs described in Examples 3-5, as well as similar plasmid constructs comprising native coding regions encoding native PA and LF, as well as empty control plasmids, are tested *in vivo* in mice by

intramuscular injection of the rectus femoris muscle within the quadriceps, using methods described above. There are 5-10 animals per group. A standard DNA vaccination protocol is used (50 µg DNA in 150 mM sodium phosphate (1 mg/ml)/leg at 0, 14 and 28 days). Alternative DNA formulations include PBS instead of sodium phosphate, adjuvants, *e.g.*, Vaxfectin™ at a 4:1 DNA: Vaxfectin™ mass ratio, mono-phosphoryl lipid A (detoxified endotoxin) from *S. minnesota* (MPL) and trehalosedicorynomycolateAF (TDM), in 2% oil (squalene)-Tween 80-water (MPL + TDM, available from Sigma/Aldrich, St. Louis, MO, (catalog # M6536)), a solubilized mono-phosphoryl lipid A formulation (AF, available from Corixa), (±)-N-(3-Acetoxypropyl)-N,N-dimethyl-2,3-bis(octyloxy)-1-propanaminium chloride (compound # VC1240), or poloxamers, *e.g.*, CRL1005 (from Organichem) and a solution of benzyl-alkonium chloride "BAK" (from Ruger Chemicals)("CRL 1005/BAK") (*see* Shriver, J.W. *et al.*, *Nature 415*:331-335 (2002), and P.C.T. Publication No. WO 02/00844 A2, each of which is incorporated herein by reference in its entirety); or transfection-facilitating cationic lipids, *e.g.*, DMRIE/DOPE at a 4:1 DNA:lipid mass ratio.

Serum samples for antibody assays are taken at 0, 21, and 41 days. On or about day 42, the vaccinated animals are challenged using either a tail vein injection of purified lethal factor toxin (Letx) or pulmonary delivery of aerosolized *B. anthracis*. Mice are challenged using the purified *B. anthracis* lethal toxin (Letx), *i.e.*, the combined mature PA65 and LF proteins. These proteins are provided through a collaborative agreement with Dr. Stephen Leppla, National Institutes of Dental Research, at the NIH. The proteins are expressed in *E.coli* as recombinant proteins and purified according to published protocols (*see*, *e.g.*, Leppla, SH *Methods Enzymol. 165*:103-116 (1988) and Park, S and Leppla, SH *Protein Expr. Purif. 18*:293-302 (2000), each of which are incorporated herein by reference in their entireties). The challenge is conducted by injecting the mouse tail vein with a protein cocktail containing 60 µg of purified PA and 25-30 µg of purified LF. This

approximates the equivalent of five 50% lethal doses of Letx. The animals are monitored for morbidity and mortality at regular intervals following challenge.

b) Rabbit immunizations.

The rabbit has increasingly gained acceptance as a relevant animal model to evaluate efficacy of vaccines against *B. anthracis*. The plasmid constructs described in Examples 3-5, as well as similar plasmid constructs comprising native coding regions encoding native PA and LF, as well as empty control plasmids, are tested *in vivo* in rabbits by the following method. Plasmid vaccination of rabbits is done at four-week intervals. At each time point, the animals (n=2-4) receive an IM injection (quadriceps) of 500 μg (1mg/ml) of DNA in 150 mM sodium phosphate formulated with the adjuvant Vaxfectin™ at a 4:1 DNA:Vaxfectin™ mass ratio, each animal receiving a total of 3 injections (1500 μg/animal). Alternative DNA formulations include other adjuvants as described herein, for example, CRL1005/BAK (*see* Shriver, J.W. *et al.*, *Nature 415*:331-335 (2002), and P.C.T. Publication No. WO 02/00844 A2), and/or transfection-facilitating cationic lipids, *e.g.*, DMRIE/DOPE at a 4:1 DNA:lipid mass ratio. Serum samples are taken at Day 0, 42, and 69 to determine antibody titers. The animals receive an aerosolized challenge on Day 70.

Rabbits are challenged in a BSL-3 facility (available, for example, at the Battelle Medical Research Evaluation Facility (MREF) in West Jefferson, OH) by standard methods. See, *e.g.*, Henderson, DW *J. Hygiene 50*:53-68 (1952)). The Battelle facility has the equipment, staff, and certification to safely conduct a aerosol challenge of large mammals using infectious and toxin producing *B. anthracis*. Vaccinated animals are transferred to Battelle's facility in West Jefferson, and then, after a IACUC approved holding isolation period, the animals are challenged with between 50 and 100 LD50 aerosolized *B. anthracis* spores by inhalation. The animals are monitored for morbidity and mortality at regular intervals following challenge.

c) Non-human primate immunizations.

The plasmid constructs described in Examples 3-5, as well as similar plasmid constructs comprising native coding regions encoding native PA and LF, as well as empty control plasmids, are tested *in vivo* in non-human primates by the following method. Cynomolgus macaques (*M. fascicularis*) are used for immunization and challenge experiments. Plasmid vaccination of the macaques is done at four-week intervals. Animals receive 1 to 1.5 mg each of DNA at each immunization bilaterally (2 to 3 mg total) intramuscularly, in the deltoid muscle. Following immunization, all animals are challenged by pulmonary delivery of aerosolized *B. anthracis*.

d) Human immunizations.

The plasmid constructs described in Examples 3-5, as well as similar plasmid constructs comprising native coding regions encoding native PA and LF, as well as empty control plasmids, are tested *in vivo* in healthy human volunteers by the following method. The plasmids are formulated in 150 mM sodium phosphate, optionally including Vaxfectin™ at a 4:1 DNA: Vaxfectin™ mass ratio, and or a poloxamer, *e.g.*, 0.01% (w/v) Pluronic® R 25R2. Vaccinations are given at 0, 4, and 8 weeks intramuscularly into the deltoid muscle either by needle injection or by needleless Biojector jet (*see, e.g.*, Wang, R. *et al. Proc. Natl. Acad. Sci.* USA *98*:10817-10822 (2001)). The volunteers receive 1 to 1.5 mg each of DNA at each immunization. Following immunization, serum specimens are collected from the volunteers and tested for antibodies to *B. anthracis* LF or PA.

e) Laboratory animal, companion animal, or food animal immunizations.

Plasmid constructs such as those described in Examples 3-5, are prepared using codon-optimized coding regions optimized for the species of

interest using an appropriate codon-usage table, *e.g.*, Table 3 (mouse), Table 4 (domestic cat), or Table 5 (cow). Codon optimization may be carried out by using relative frequencies for the codons, or by using the most frequent codon, as described herein. Plasmids comprising these coding regions, plus similar plasmid constructs comprising native coding regions encoding native PA and LF, as well as empty control plasmids, are tested *in vivo* in various animal species by the following method. The animal species of interest is immunized with an appropriate amount of a DNA vaccine codon-optimized for that species, at an appropriate amount, delivered in an appropriate route for that species, including, but not limited to the following immunization strategies: for mouse immunization, intramuscular delivery into the rectus femoris muscle of 50 μg DNA in 150 mM sodium phosphate (1 mg/ml)/leg at 0, 14 and 28 days; for cow immunization, intradermal delivery into the ear of 500 μg DNA in normal saline (1 mg/ml) at days 0 and 21 (*see, e.g*, van Drunen Little-van den Hurk *et al. J. Gen. Virol. 79*:831-839 (1998)); and for domestic cat immunization, intradermal delivery of 300 μg DNA in normal saline (1 mg/ml) at days 0, 15, and 30 (*see, e.g.*, Osorio, JE, *et al. Vaccine 17*:1109-1116 (1999)).

EXAMPLE 9

Immunological Asays

a) ELISA for LF and PA Antibody Titers.

Microtiter plates are coated with either PA or LF antigen by incubating 100 ng/well of purified protein (obtained from List Biological Laboratories, Campbell, CA) overnight at 4°C in 100 mM carbonate buffer, pH 9.6. The wells are washed (3X) with 10 mM Tris-buffered (pH 7.3); 150 mM NaCl (TBS) followed by a 1% (w/v) BSA block. Serially diluted experimental and control serum samples in TBS + 0.05% Tween are added to the wells and incubated for 60 min at room temperature. Enzyme conjugated (horseradish

peroxidase or alkaline phosphatase) anti-mouse, anti-rabbit, or anti-monkey IgG are then added to each well and supernatants monitored for enzyme product. Antibody titers are defined as the highest dilution of a serum sample that results in an absorbance value 2X greater than that of a non-immune control serum. Antibody quantification will be determined using a purified anti-PA and anti-LF IgG1 and IgG2 reagent antibody.

b) Toxin Neutralization Assay.

Antibodies from vaccinated animals are initially tested using an *in vitro* assay that measures the neutralization of lethal toxin (Letx, *i.e.*, LF and PA protein) cytotoxicity. Briefly, this protection assay is carried out using 24 hr. cultures of J774A.1 mouse macrophage cells maintained in microtiter plates (~6 X 10 ^{4}cells per well) in DME media, with glucose and L glutamine supplements, and 7% fetal bovine serum at 37°C. Serially diluted serum from vaccinated and control animals are mixed with letx and allowed to sit for 60 min. The final Letx concentration will be brought to 3 μg/ml. This mixture will then added to the J774A.1 cells and incubated for seven hours at 37°C. Finally, 100 μl of 0.5 mg/ml 3-[4,5-dimethylthiazol-2yl]-2,5-diphenyltetrazolium bromide (MTT) is added to each experimental well and allowed to incubate another 60 min. before assaying for cytotoxicity. In this assay, surviving cells metabolize MTT into an insoluble purple pigment in a manner that is proportional to viability. This insoluble pigment is recovered from viable cells and quantitated by absorption of 450nm light.

EXAMPLE 10

Immunization using a Prime-Boost Strategy

There is accumulating evidence to suggest that a naked DNA prime with a heterologous viral or protein boost will result in an enhanced humoral response. Since the humoral response is widely believed to be the immune

correlate of protection against *B. anthracis*, in certain experiments a prime-boost strategy is used. The boost may be purified non-toxic LF and/or PA protein or the commercially available AVA vaccine. Alternatively recombinant virus vectors, *e.g.*, adenovirus vectors, expressing non-toxic LF and/or PA may be used as the boost. Results are evaluated to compare antibody titers resulting from prime-boost immunization relative to DNA vaccination alone.

New Zealand rabbits are immunized with a series of three plasmid injections or two plasmid injections with the plasmid constructs described in Examples 3-5, as well as similar plasmid constructs comprising native coding regions encoding native PA and LF, as well as empty control plasmids, followed by a single dose of recombinant PA and/or LF protein (1 microgram in Alhydrogel) or the AVA vaccine (5 microliters). Controls include immunization with the codon optimized and control plasmid constructs alone, and mock immunizations. Following the immunization series, *e.g.*, two plasmid DNA immunizations at four week intervals followed by a boost at week 12, total antibody titers and neutralizing titers are determined. In addition, selected immunized animals are challenged with a 500x LD50 dose of aerosolized anthrax spores at Battelle Medical Research Evaluation Facility in West Jefferson, OH as described in Example 8.

EXAMPLE 11

Immunization of Mice Using Codon-Optimized *B. anthracis* DNA Vaccines

a) Experiment 1

Six groups (Groups A-F) of 5 Balb/c female mice were injected bilaterally in the rectus femoris muscle with 50 µl of DNA solution (at 1.0 mg/ml) (100 µl total/mouse), on days 1 and 21 and 42 with each of the following plasmids:

Group 1A: VR6290 (TPA-PA63, Fig. 1, SEQ ID NO:1, prepared as described in Example 3a);

Group 1B: VR6291 (TPA-PA63ΔFF, Fig. 2, SEQ ID NO:5, prepared as described in Example 3b);

Group 1C: VR6292 (TPA-PA83Δfurin, Fig. 3, SEQ ID NO:7, prepared as described in Example 3c);

Group 1D: VR6295 (TPA-LF HEXXH, Fig. 4, SEQ ID NO:9, prepared as described in Example 4a);

Group 1E: VR6290 (50 μg) +VR6295 (50 μg), co-injected; and

Group 1F: VR1012 (empty expression vector).

The plasmids listed above were forumulated as follows. One vial (0.5 mg) of MPL+TDM adjuvant, purchased from Sigma/Aldrich (catalog # M6536) was resuspended in 150 mM Na_2PO_4 according to manufacturers instructions. Fifty microliters of DNA solution was mixed 1:1 (v/v) with the MPL+TDM emulsion and injected into each mouse at the times specified above.

Mice were bled for serum on days 0 (prebleed) , 20 (bleed 1), and 41 (bleed 2), and 62 (bleed 3). PA antibodies were measured in each of Groups 1A-1C, 1E, and 1F, LF antibodies were measured in each of Groups 1D, 1E, and 1F, and LT neutralizing antibodies were measured in each of Groups 1A-1E. All assays were done as outlined in Example 9. The geometric mean of the anti-PA and anti-LF titers were calculated following each bleed. The results are shown in Figs. 15A and 15B, respectively. In Fig. 15C, the serum from each mouse was tested for LT neutralizing antibody titer after the last DNA immunization (bleed 3) according to the procedure in Example 9. The mean neutralizing titer for each group of mice was calculated and plotted and the error bars represent one standard deviation from the mean.

b) Experiment 2

Eight groups of 5 mice each (Groups 2A-2H) were injected bilaterally in the rectus femoris with 50 μl (50 μg) of DNA solution (100 μl (100 μg)

total per mouse), adjuvanted with MPL + TDM as described in section 11a, on days 1, 21, and 49 with the following combinations of plasmids:

Group 2A: VR-6290 (50 μg) + VR-1012 (50 μg);

Group 2B: VR-6291 (50 μg) + VR-1012 (50 μg);

Group 2C: VR-6292 (50 μg) + VR-1012 (50 μg);

Group 2D: VR-6295 (50 μg) + VR-1012 (50 μg);

Group 2E: VR-6290 (50 μg) + 50 μg VR-6295

Group 2F: VR-6291 (50 μg) + VR-6295 (50 μg);

Group 2G: VR-6292 (50 μg) + VR-6295 (50 μg); and

Group 2H: VR-1012 (100 μg).

Mice were bled for serum on days 0 (prebleed) , 20 (bleed 1), and 41 (bleed 2), and 62 (bleed 3). PA antibodies were measured in each of Groups 2A-2C and 2E-2H, LF antibodies were measured in each of Groups 2D-2H, and LT neutralizing antibodies were measured in each of Groups 2A-2G. All assays were done as outlined in Example 9. The geometric mean of the anti-PA and anti-LF titers were calculated following each bleed. The results are shown in Figs. 16A and 16B, respectively. In Fig. 16C, the serum from each mouse was tested for LT neutralizing antibody titer after the last DNA immunization (bleed 3) according to the procedure in Example 9. The mean neutralizing titer for each group of mice was calculated and plotted and the error bars represent one standard deviation from the mean.

c) Experiment 3

Four groups of 5 mice each (Groups 3A-3D) were injected bilaterally in the rectus femoris with 50 μl (50 μg) of DNA solution (100 μl (100 μg) total per mouse), adjuvanted with MPL + TDM as described in section 11a, on days 1, 21, and 49 with the following combinations of plasmids:

Group 3A: VR-6292 (50 μg) + VR-1012 (50 μg);

Group 3B: VR-6292 (50 µg) + VR-62952 (50 µg, TPA-LF Domain I-III, Fig. 5, SEQ ID NO:13, prepared as described in Example 4b);

Group 3C: VR-6292 (50 µg) + VR-62951 (50 µg, TPA-LF Domain IB, Fig. 14, SEQ ID NO:39, prepared as described in Example 4d); and

Group 3D: VR-6299 (50 µg, TPA-Sugar minus PA63, Figure 7, SEQ ID NO:17, prepared as described in Example 5a) + VR-1012 (50 µg).

Mice were bled for serum on days 0 (prebleed) , 20 (bleed 1), and 41 (bleed 2), and 62 (bleed 3). PA antibodies were measured in each of Groups 3A-3D, LF antibodies were measured in each of Groups 3B and 3C, and LT neutralizing antibodies were measured in each of Groups 3A-3D. All assays were done as outlined in Example 9. The geometric mean of the anti-PA and anti-LF titers were calculated following each bleed. The results are shown in Figs. 17A and 17B, respectively. In Fig. 17C, the serum from each mouse was tested for LT neutralizing antibody titer after the last DNA immunization (bleed 3) according to the procedure in Example 9. The mean neutralizing titer for each group of mice was calculated and plotted and the error bars represent one standard deviation from the mean.

d) Experiment 4

Four groups of 10 mice each (Groups 4A-4D) were injected bilaterally in the rectus femoris with 50 µl (50 µg) of of plasmid VR-6292 (100 µl (100 µg) total per mouse), formulated with various adjuvants, on days 1, 21, and 49, as follows:

Group 4A: VR-6292 formulated with CRL 1005/BAK;

Group 4B: VR-6292 formulated with MPL + TDM, as described in section 11a, *supra*;

Group 4C: VR-6292 formulated with Vaxfectin™ at a 4:1 DNA: Vaxfectin™ mass ratio; and

Group 4D: VR-6292 formulated with DMRIE:DOPE (1:1 molar ratio) at a 4:1 DNA:lipid mass ratio.

The plasmids in Group 4A were formulated as follows. The poloxamer CRL1005 (from Organichem) and a solution of benzyl-alkonium chloride "BAK" (from Ruger Chemicals) were added sequentially to plasmid solutions in PBS. The initial plasmid/poloxamer formulation was prepared to contain 5 mg/mL plasmid DNA, 7.5 mg/mL CRL1005 and 0.3 mM BAK. These initial preparations were diluted 1:1 (vol:vol) with PBS, then cold sterile filtered. Further dilution with sterile PBS was done just prior to use to provide the final working concentration of 1 mg/mL pDNA, 1.5 mg/mL CRL1005 and 0.06 mM BAK.

Mice were bled for serum on days 0 (prebleed) , 20 (bleed 1), and 41 (bleed 2), and 62 (bleed 3). PA antibodies were measured in each of Groups 4A-4D, and LT neutralizing antibodies were measured in each of Groups 4A-4D. All assays were done as outlined in Example 9. The geometric mean of the anti-PA titers were calculated following each bleed. The results are shown in Fig. 18A. In Fig. 18B, the serum from each mouse was tested for LT neutralizing antibody titer after the last DNA immunization (bleed 3) according to the procedure in Example 9. The mean neutralizing titer for each group of mice was calculated and plotted and the error bars represent one standard deviation from the mean.

e) Experiment 5

Six groups of 10 mice each (Groups 5A-5F) were injected bilaterally in the rectus femoris with 50 µl (50 µg) of of plasmid VR-6292 (100 µl (100 µg) total per mouse), formulated with various adjuvants, on days 1, 14, and 28, as follows:

Group 5A: VR-6292 formulated with MPL + TDM, as described in section 11a, *supra*;

Group 5B: VR-6292 formulated with MPL-A aqueous 1000 µg/mL (Corixa) mixed 1:1 (v/v) with DNA;

Group 5C: VR-6292 formulated with CRL 1005/BAK, as described in section 11d, *supra*;

Group 5D: VR-6292 formulated with Vaxfectin™ at a 4:1 DNA: Vaxfectin™ mass ratio;

Group 5E: VR-6292 formulated with DMRIE:DOPE (1:1 molar ratio) at a 4:1 DNA:lipid mass ratio; and

Group 5F: VR-6292 formulated with 1 X PBS.

Mice were bled for serum on days 0 (prebleed) , 13 (bleed 1), and 27 (bleed 2), and 56 (bleed 3). PA antibodies were measured in each of Groups 5A-5F after each bleed, and LT neutralizing antibodies were measured in each of Groups 5A-5F after bleed 3. All assays were done as outlined in Example 9. The geometric mean of the anti-PA titers were calculated following each bleed. The results are shown in Fig. 20. In Fig. 21, the serum from each mouse was tested for LT neutralizing antibody titer after the last DNA immunization (bleed 3) according to the procedure in Example 9. The mean neutralizing titer for each group of mice was calculated and plotted and the error bars represent one standard deviation from the mean.

EXAMPLE 12

Immunization of Rabbits Using Codon-Optimized *B. anthracis* DNA Vaccines

Twelve (12) groups of 10 rabbits each (Groups A-G and I-M, for DNA vaccinations) and one group of 4 rabbits (Group H, for the AVA vaccination) (*Oryctolagus cuniculus*, New Zealand albino rabbits, 2-5 kg each at onset of treatment) were used in this experiement. The rabbits in Groups A-G and I-M received a 500 μg intramuscular injection in each quadricep muscle (bilateral) for a total of 1 mg of plasmid DNA per rabbit per immunization. Injection of the formulated plasmid DNA took place on days 0, 28, and 56. Some animals received only the first two plasmids injections on days 0 and 28 (denoted 2 injs in Fig. 19). All rabbits were prebled two days before the first immunization and bled again on days 14, 42, and 70.

Unless noted, the various formulations were administered by a bilateral intramuscular injection into the quadriceps muscles on Days 0, 28, and 56 with a needle. The dose volume to be administered is 500 μl/muscle, 1 ml/animal. The rabbits in Group D were vaccinated using a Biojector, as follows. Animals were anesthetized using ketamine/xylazine. The skin over the injection site was shaved, and the dose volume administered was 500 μl/muscle, 1 ml/animal. The vaccination groups were as follows:

Group A: VR6292 formulated with Vaxfectin™ at a 4:1 DNA: Vaxfectin™ ratio;

Group B: VR6292 (500 μg) + VR-62952 (500 μg) formulated with Vaxfectin™ at a 4:1 DNA: Vaxfectin™ ratio;

Group C: VR6292 formulated with DMRIE/DOPE at a 4:1 DNA:lipid ratio;

Group D: VR6292 formulated with Vaxfectin™ at a 4:1 DNA: Vaxfectin™ ratio, delivered by Biojector;

Group E: VR6292 (500 μg) + VR-62951 (500 μg) formulated with Vaxfectin™ at a 4:1 DNA: Vaxfectin™ ratio;

Group F: VR6290 formulated with Vaxfectin™ at a 4:1 DNA: Vaxfectin™ ratio;

Group G: VR6292 formulated with Vaxfectin™ at a 4:1 DNA: Vaxfectin™ ratio (two injections only);

Group H: Commercial anthrax vaccine AVA, 50 μl, delivered on day 28 and 56 by a single IM injection;

Group I: VR-62951 formulated with Vaxfectin™ at a 4:1 DNA: Vaxfectin™ ratio;

Group J: VR6292 (500 μg) + VR-62951 (500 μg) formulated with Vaxfectin™ at a 4:1 DNA: Vaxfectin™ ratio (two injections only);

Group K: VR-62952 formulated with Vaxfectin™ at a 4:1 DNA: Vaxfectin™ ratio;

Group L: VR6292 formulated with MPL-A aqueous 1000 μg/mL (Corixa) mixed 1:1 (v/v) with DNA;

Group M: VR6292 formulated with CRL1005/BAK, formulated as described in Example 11d, *supra*.

The LT neutralization assay was performed on all rabbit sera from the day 70 bleed. The median titer ± one standard deviation is shown for each group in Fig. 19.

EXAMPLE 13

Immunization and Challenge of Rabbits Using Codon-Optimized *B. anthracis* DNA Vaccines

Ten groups of rabbits (*Oryctolagus cuniculus*, New Zealand albino rabbits, 2-5 kg each at onset of treatment, ten (10) animals per group unless otherwise noted) were used in this experiment. These included selected groups of animals described in Example 12, as noted below. The various plasmid DNA formulations were administered by a bilateral intramuscular injection into the quadriceps muscles on Days 0, 28, and 56 with a needle. The dose volume to be administered is 500 µl/muscle, 1 ml/animal. The vaccination groups were as follows:

Group 1: VR6292 formulated with Vaxfectin™ at a 4:1 DNA: Vaxfectin™ ratio (Group A from Example 12);

Group 2: VR6292 formulated with Vaxfectin™ at a 4:1 DNA: Vaxfectin™ ratio (two injections only) (Group G from Example 12);

Group 3: VR6292 formulated with DMRIE/DOPE at a 4:1 DNA:lipid ratio (Group C from Example 12);

Group 4: VR6292 (500 µg) + VR-62951 (500 µg) formulated with Vaxfectin™ at a 4:1 DNA: Vaxfectin™ ratio (two injections only) (Group J from Example 12);

Group 5: VR6292 (500 µg) + VR-62952 (500 µg) formulated with Vaxfectin™ at a 4:1 DNA: Vaxfectin™ ratio (Group B from Example 12);

Group 6: VR-62952 formulated with Vaxfectin™ at a 4:1 DNA: Vaxfectin™ ratio (two animals) (Group K from Example 12);

Group 7: VR1012 formulated with Vaxfectin™ at a 4:1 DNA: Vaxfectin™ ratio (four animals);

Group 8: VR1012 formulated with DMRIE/DOPE at a 4:1 DNA:lipid ratio (two animals);

Group 9: Commercial anthrax vaccine AVA, 50 µl, delivered on 28 and 56 by a single IM injection (Group I from Example 12); and

Group 10: Twelve unvaccinated animals.

The rabbits in Groups 1-8 received a 500 µg intramuscular injection in each quadricep muscle (bilateral) for a total of 1 mg of plasmid DNA per rabbit per immunization. In groups 1, 3, 5, 6, 7, 8, and 9, three injections of the formulated plasmid DNA took place on days 0, 28, and 56. In groups 2 and 4, two injections of the formulated plasmid DNA took place on days 0 and 28. In group 19 commercial anthrax vaccine AVA, 50 µl, was injected intramuscularly on days 28 and 56. All rabbits were prebled two days before the first immunization and bled again on days 14, 42, and 70.

Over a four-day period on or around day 70 (indicated in Table 17 as "challenge days" C1-C4), the rabbits were challenged by aerosol administration of *B. anthracis* (Ames strain) spores by standard methods. *See, e.g.*, Henderson, DW *J. Hygiene 50*:53-68 (1952)). Challenge doses ranged from about 50 LD50 equivalents to about 250 LD50 equivalents as noted in Table 17 below. The animals were monitored for morbidity and mortality at regular intervals out to days 19-22 (depending on the challenge day) following challenge. The results are shown in Table 17, and are summarized in Table 18. "NC" denotes "not challenged."

TABLE 17

ANIMAL ID	CHALLENGE DAY	AMES LD50 EQUIVALENT	SURVIVAL
Group 1			
1.1	C3	123.5	Y
1.2	NC		
1.3	C4	113.5	Y
1.4	C1	56.4	Y
1.5	C3	92.7	Y
1.6	C4	66.3	Y
1.7	C1	103	Y
1.8	C2	127.3	Y
1.9	NC		
1.10	C2	128.8	Y
Group 2			
2.1	C1	76	Y
2.2	C4	70.5	Y
2.3	NC		
2.4	C4	52.1	Y
2.5	C1	252.1	Y
2.6	C2	119.1	Y
2.7	C3	52.4	Y
2.8	NC		
2.9	C2	71.9	Y
2.10	C3	195.1	Y
Group 3			
3.1	C2	55.7	Y
3.2	C3	238.3	Y
3.3	C4	110	Y
3.4	C1	208.1	Y
3.5	NC		
3.6	C1	142.9	Y
3.7	C3	169	Y
3.8	NC		
3.9	C4	57.5	Y
3.10	C2	74.7	Y
Group 4			
4.1	C3	87.3	Y
4.2	C4	90.2	Y
4.3	C1	81.6	Y
4.4	NC		
4.5	C2	100	Y
4.6	C2	72	Y
4.7	C1	76.1	Y
4.8	C4	92.8	Y
4.9	NC		
4.10	C3	205	Y

Group 5			
5.1	NC		
5.2	C1	192.2	Y
5.3	C2	152.6	Y
5.4	C4	66.6	Y
5.5	C3	135.7	Y
5.6	C2	65.1	Y
5.7	C4	79	Y
5.8	C1	126.6	Y
5.9	C3	154.7	Y
5.10	NC		
Group 6			
6.1	C4	117.7	Y
6.2	C3	241.4	N (D4)
6.3	NC		
6.4	C1	107.3	Y
6.5	C4	58.7	N (D4)
6.6	C3	121	Y
6.7	C3	160.8	Y
6.8	C2	46.1	N (D7)
6.9	C1	195.2	N (D6)
6.10	C2	94.5	Y
Group 7			
7.1	C3	101.9	N (D3)
7.2	C4	144.1	N (D2)
7.3	NC		
7.4	C1	108.2	N (D2)
Group 8			
8.1	C2	63	N (D3)
8.2	C4	58.2	N (D3)
Group 9			
9.1	C2	113.4	Y
9.2	C1	106.9	Y
9.3	C3	157.6	Y
9.4	C4	175.6	Y
Group 10			
10.1	C4	76.7	N (D2)
10.2	C3	207.6	N (D3)
10.3	C2	91.5	N (D2)
10.4	C4	176	N (D2)
10.5	C2	123	N (D3)
10.6	C2	95.4	N (D3)
10.7	C4	91.5	N (D3)
10.8	C1	165.2	N (D2)
10.9	C1	57.3	N (D3)
10.10	C3	163.8	N (D4)
10.11	C3	114.2	N (D2)
10.12	C1	62.3	N (D2)

TABLE 18

Group	Survival
1	8/8 (100%)
2	8/8 (100%)
3	8/8 (100%)
4	8/8 (100%)
5	8/8 (100%)
6	5/9 (56%)
7	0/2 (0%)
8	0/3
9	4/4 (100%)
10	0/12 (0%)

EXAMPLE 14

Immunization of Mice using Single Vial Formulations

Single vial formulations were prepared by reconstituting bulk DMRIE and DOPE lipids to form multi-lamellar vesicles (MLV). These vesicles were then further processed to produce small DMRIE and DOPE liposomes (SUV) and sterile filtered through a 0.2μm membrane. The formulations were prepared aseptically at room temperature by adding sterile plasmid DNA and sterile DMRIE:DOPE SUV liposomes into separate feed lines and then combining into a third sterile vessel via in-line mixing. Moderate rates of addition and moderate in-vessel mixing were used to form a lipid/plasmid DNA complex. Preparation of lipids and lipid/plasmid DNA complexes is described in Zelphati *et al. Gene Therapy 5*: 1277-1282 (1998) which is incorporated herein by reference in its entirety. The formulations described below contain final molar ratios of 4:1 or 2:1 plasmid DNA to DMRIE.

Eight groups of mice, containing 10 mice in each group, were injected bilaterally in the rectus femoris muscle with the various formulations described below. Each injection contained 50μg of purified plasmid VR6292 (PA83Δfurin) in a volume of 0.1ml. At 0, 2 and 4 weeks the groups were injected with the following formulations, all containing 50μg of VR6292

plasmid DNA (prepared as described in the plasmid DNA purification section prior to Example 1).

Group A: Unextruded MLV, in a 4:1 molar ratio of plasmid DNA to DMRIE, in PBS (pH 7.2). The formulation was freshly prepared just prior to injection.

Group B: Unextruded MLV, in a 4:1 molar ratio of plasmid DNA to DMRIE, in 10% sucrose and 10mM sodium phosphate, pH 7.2. The formulation was freshly prepared just prior to injection.

Group C: 0.2 µm filter extruded (SUV) liposomes, in a 4:1 molar ratio of plasmid DNA to DMRIE, in 10% sucrose and 10mM sodium phosphate (pH 7.2). The formulation was stored overnight at 2-8°C prior to inoculation.

Group D: SUV liposomes, in a 4:1 ratio plasmid DNA to DMRIE, in 10% sucrose and 10mM sodium phosphate, pH 7.2. The formulation was frozen prior to inoculation.

Group E: SUV liposomes, in a 4:1 molar ratio of plasmid DNA to DMRIE, in 10% sucrose and 10mM sodium phosphate, pH 7.2. The formulation was lyophilized prior to inoculation.

Group F: Unextruded MLV, containing cholesterol in place of DOPE, in a 4:1 molar ratio of plasmid DNA to DMRIE, in 10% sucrose and 10mM sodium phosphate, pH 7.2. The formulation was freshly prepared just prior to injection.

Group G: Unextruded MLV, in a 2:1 molar ratio of plasmid DNA to DMRIE, in PBS, pH 7.2. The formulation was freshly prepared just prior to injection.

Group H: SUV liposomes, in a 2:1 molar ratio of plasmid DNA to DMRIE, in 10% sucrose and 10mM sodium phosphate, pH 7.2. The formulation was stored overnight at 2-8°C prior to injection.

Mice were bled for serum prior to each DNA immunization at week 0 (Prebleed), week 2 (Bleed 1), week 4 (Bleed 2) and four weeks post the last injection (Bleed 3). Anti-PA IgG antibody titers and neutralization of lethal toxin (Letx) titers were performed as described in Example 9. The antibody titers and neutralization results for each bleed and every formulation tested are shown in Tables 19 and 20.

TABLE 19: Anti – PA IgG Titer

Group		A	B	C	D
Geometric Mean	Prebleed	80	80	80	80
	Bleed 1	10975	4165	7760	4457
	Bleed 2	305736	66540	108094	62084
	Bleed 3	1616014	376405	376405	655627
Std. Dev.	Prebleed	0	0	0	0
	Bleed 1	11372	7630	23983	12421
	Bleed 2	215705	58765	84998	51642
	Bleed 3	639310	370406	343674	1200361

Group		E	F	G	H
Geometric Mean	Prebleed	80	80	80	80
	Bleed 1	7760	5487	9554	4457
	Bleed 2	81920	71316	327680	76434
	Bleed 3	1310720	266159	1310987	351199
Std. Dev.	Prebleed	0	0	0	0
	Bleed 1	15017	8172	23498	26219
	Bleed 2	116014	55555	205073	53970
	Bleed 3	0	197337	678738	221840

TABLE 20: Letx Neutralizing Antibody Titers

Group	A	B	C	D
Mean	184	226	160	149
Std. Dev.	217	372	178	390

Group	E	F	G	H
Mean	92	86	211	35
Std. Dev	111	89	212	46

EXAMPLE 15

Immunization of Non-Human Primates Using Codon-Optimized *B. anthracis* DNA Vaccines

Three groups of cynomologous macaques (*M. fasicularis*), containing three monkeys in each group, were used in this experiment. The animals were immunized unilaterally, intramuscularly, in the deltoid muscle with a Bioinjector device. Varying amounts of purified VR6292 (PA83Δfurin) plasmid DNA (prepared as described in the plasmid DNA purification section prior to Example 1) formulated with Vaxfectin™, in a 4:1 molar ratio of plasmid DNA to lipid, was used in all inoculations in this study. All animals received injections at month 0, 1 month, and 2 months. Group 1 received 20μg of plasmid DNA at each inoculation. Group 2 received 100μg of plasmid DNA at each inoculation. Group 3 received 200μg of plasmid DNA at each inoculation.

The monkeys were bled for serum prior to each DNA immunization at month 0 (Bleed 1), month 1 (Bleed 2), month 2 (Bleed 3) and at four weeks after the last injection (Bleed 4). Anti-PA IgG antibody titers and neutralization of lethal toxin (Letx) titers were performed as described in Example 9. The antibody titers and neutralization results for each group of animals are shown in Tables 21, 22 and 23.

2 out of 3 animals in Group 1 generated an anti-PA IgG titer. One of the animals generated a sizable titer (20,000) after three injections. This titer is comparable to the titers of the animals in groups receiving higher doses of plasmid DNA (Groups 2 and 3). None of the animals in Group 1 generated any measurable neutralization activity of Letx at the lowest dilutions tested (serum diluted 1:20).

The animals in Groups 2 and 3 generated similar immune responses to the inoculations. All monkeys in both groups developed anti-PA IgG titers. Letx neutralization titers were generated in 2 out of 3 monkeys in both groups. The remaining animal in each group had measurable neutralization activity, but below the level needed to score a titer.

TABLE 21: Group 1 – Anti-PA IgG and LetX Neutralizing Titers

Animal #		**1001**	**1002**	**1003**
	Bleed 2	80	160	640
	Bleed 3	80	640	10240
	Bleed 4		2560	20480
	Letx Neutralizing Titer			
	Bleed 3	0	0	0
	Bleed 4	0	0	0

TABLE 22: Group 2 - Anti-PA IgG and LetX Neutralizing Titers

	Anti-PA IgG			
Animal #		**2001**	**2002**	**2003**
	Bleed 2	640	10240	5120
	Bleed 3	10240	20480	40960
	Bleed 4	40960	40960	81920
	Letx Neutralization Titer			
	Bleed 3	0	***	***
	Bleed 4	***	40	80

*** denotes a detectable low level of neutralization activity

TABLE 23: Group 3 - Anti-PA IgG and LetX Neutralizing Titers

	Anti-PA IgG			
Animal #		**3001**	**3002**	**3003**
	Bleed 2	640	320	640
	Bleed 3	5120	20480	10240
	Bleed 4	20480	20480	81920
	Letx Neutralizing Titer			
	Bleed 3	0	40	***
	Bleed 4	***	160	80

*** denotes a detectable low level of neutralization activity

EXAMPLE 16

Immunization Challenge of Rabbits Using Codon-Optimized *B. anthracis* DNA Vaccines

a) Long-Term Immune Response in DNA Immunized Rabbits

10 rabbits immunized, as described in Example 12, Group D (Immunized three times with VR6292), were followed long-term for anti-PA antibody titer, LetX neutralization titer and protective immune response to an anthrax spore challenge. Anti-PA IgG antibody titers and LetX neutralization titers were performed as described in Example 9. The results of the titers and neutralization assays are shown in Table 24. Rabbits were bled twelve times on the weeks indicated in Table 24.

On week 39 of the experiment, rabbits were challenged by aerosol administration of *B. anthracis* (Ames strain) spores by standard methods as described in Example 13. All rabbits survived. Control animals that were not vaccinated did not survive challenge.

TABLE 24

Anti-PA IgG Serum Antibody Titer		
Week (post first injection)	Geometric Mean	Std. Dev.
2	20480	22744
6	2622775	1635799
10	12909485	6079129
14	6456057	8244196
18	4565122	7496131
22	2810448	2931335
26	1311120	1508050
30	1311120	1508050
34	1405367	888676
39	1064744	1475391
40	1505928	1809833
42	2129704	1865167
Letx Neutralization Titer		
Week (post first injection)	Geometric Mean	Std. Dev.
6	1576	843
10	4457	2956
14	2560	1602
18	1372	1441
22	1194	1455
26	970	607
30	905	641
35	905	641
40	844	843
41	1040	911
43	1194	955

b) Rabbit Immunization Dosing with Intended Human Vaccine Product

Sixty New Zealand White rabbits (30 males and 30 females), approximately 10-12 weeks old, were used for this study. Ten animals per sex were injected with the formulations described below. The plasmids were formulated with DMRIE/DOPE in a 4:1 DNA to lipid mass ratio in PBS, as described in Example 8b.

Group 1: 1.0ml of PBS

Group 2: 0.1mg of plasmid VR-6292 and 0.1mg of

plasmid VR 62952.

Group 3: 1.0mg of plasmids VR-6292 and 1.0mg VR-62952.

All animals in the study received unilateral intramuscular injections into the vastus lateralus muscle at 0, 2, 4 and 8 weeks.

Serum samples were taken from all study animals once during the pre-treatment period and once during weeks 2, 4, 6, 8, 10 and 12. Anti-PA and LF antibody titers and Letx neutralizing antibody titers were evaluated using serum samples taken prior to immunization and at 8 weeks prior to the fourth DNA immunization. All immunological assays were performed as described in Example 9. Anti-PF and LF antibody titers and Letx neutralizing antibody results for the bleeds taken at week 8 are shown in Tables 25 and 26.

TABLE 25: Anti-PA and Anti-LF Antibody Titers (Geometric Mean)

Group	**2**	**3**
Anti-PA	163840	514211
Anti-LF	163840	678540
	Std. Dev.	
Anti-PA	230686	595754
Anti-LF	386254	730464

TABLE 26: Letx Neutralization Titers (Geometric Mean)

Group	2	3
Letx titer	889	2840
Std Dev	510	1518

c) Post Challenge Immune Responses in Aerosolized Spore Challenged Rabbits.

Six groups of rabbits, with 10 individuals in each group, were immunized as described for Groups A-C, G, H and K in Example 12. 39 weeks after the last immunization, these rabbits were challenged with anthrax aerosolized spores, as described in Example 13. Control animals that had not been immunized were also challenged as described in Example 13. No control animal survived the challenge.

At one day prior to challenge, and at 7 and 21 days post challenge, animals were bled for serum. Anti-PA and LF IgG antibody and LetX neutralizing titers were performed as described in Example 9. It should be noted that except as described below, immunized animals had developed protective immunity since they survived challenge.

The immune responses post challenge could be divided into two groups: rabbits that showed no increase in immune response after challenge (lack of boosting) and rabbits that were boosted in their response to PA and/or LF after spore challenge.

All rabbits immunized as decribed for Groups A-C, in Example 12 (immunized with VR6292 or VR6292+VR62952, three times), demonstrated a lack of boosting. Two rabbits immunized as described for Group G, in Example 12 (immunized with VR6292 twice), had the lowest anti-PA titers pre-challenge and demonstrated a small post-challenge boost in anti-PA titer and the generation of an anti-LF response.

Several rabbits immunized as described for Group K, in Example 12 (immunized with VR62952 (LF[I-III])), did not survive anthrax spore challenge. The five surviving rabbits all had significant anti-PA titers post challenge. Additionally rabbits immunized as described for Group H in Example 12 (immunized twice with the commercial anthrax vaccine AVA), had no measurable anti-LF response pre-challenge. After challenge all rabbits

showed a boosted anti-PA titer and the generation of a strong anti-LF response.

In summary, all rabbits immunized two or three times with plasmids encoding PA or PA+LF generated strong immune responses and were able to survive anthrax spore challenge. Almost all of these rabbits showed a lack of immune response boosting post-challenge, which is consistent with sterilizing immunity. In contrast, rabbits immunized twice with 50µl of AVA exhibited a strong anti-LF response and a boosted anti-PA titer.

EXAMPLE 17

Mucosal Vaccination and Electrically Assisted Plasmid Delivery

a) Mucosal DNA Vaccination

Plasmid constructs comprising codon-optimized and non-codon-optimized coding regions encoding LF, PA or various fragments, variants or derivatives, as described herein, are delivered to BALB/c mice at 0, 2 and 4 weeks via i.m., intranasal (i.n.), intravenous (i.v.), intravaginal (i.vag.), intrarectal (i.r.) or oral routes. The DNA is delivered unformulated or formulated with the cationic lipids DMRIE/DOPE (DD), DMRIE/Cholesterol or Vaxfectin™. Serum IgG titers against the various LF and PA antigens are measured as described in Example 9, as well as Letx neutralization titers.

b) Electrically-assisted plasmid delivery

In vivo gene delivery may be enhanced through the application of brief electrical pulses to injected tissues, a procedure referred to herein as electrically-assisted plasmid delivery. See, *e.g.*, Aihara, H. & Miyazaki, J. *Nat. Biotechnol. 16:*867-70 (1998); Mir, L.M. *et al.*, *Proc. Natl Acad. Sci.* USA *96:*4262-67 (1999); Hartikka, J. *et al.*, *Mol. Ther. 4:*407-15 (2001); and Mir, L.M. *et al.*; Rizzuto, G. *et al.*, *Hum Gene Ther 11:*1891-900 (2000);

Widera, G. *et al*, *J. of Immuno. 164*: 4635-4640 (2000). The use of electrical pulses for cell electropermeabilization has been used to introduce foreign DNA into prokaryotic and eukaryotic cells *in vitro*. Cell permeabilization can also be achieved locally, *in vivo*, using electrodes and optimal electrical parameters that are compatible with cell survival.

The electroporation procedure can be performed with various electroporation devices. These devices include external plate type electrodes or invasive needle/rod electrodes and can possess two electrodes or multiple electrodes placed in an array. Distances between the plate or needle electrodes can vary depending upon the number of electrodes, size of target area and treatment subject.

The TriGrid needle array, used in examples described herein, is a three electrode array comprising three elongate electrodes in the approximate shape of a geometric triangle. Needle arrays may include single, double, three, four, five, six or more needles arranged in various array formations. The electrodes are connected through conductive cables to a high voltage switching device that is connected to a power supply.

The electrode array is placed into the muscle tissue, around the site of nucleic acid injection, to a depth of approximately 3 mm to 3 cm. The depth of insertion varies depending upon the target tissue and size of patient receiving electroporation. After injection of foreign nucleic acid, such as plasmid DNA, and a period of time sufficient for distribution of the nucleic acid, square wave electrical pulses are applied to the tissue. The amplitude of each pulse ranges from about 100 volts to about 1500 volts, *e.g.*, about 100 volts, about 200 volts, about 300 volts, about 400 volts, about 500 volts, about 600 volts, about 700 volts, about 800 volts, about 900 volts, about 1000 volts, about 1100 volts, about 1200 volts, about 1300 volts, about 1400 volts, or about 1500 volts or about 1-1.5kV/cm, based on the spacing between electrodes. Each pulse has a duration of about 1μs to about 1000μs, *e.g.*, about 1μs, about 10μs, about 50μs, about 100μs, about 200μs, about 300μs, about 400μs, about 500μs, about 600μs, about 700μs, about 800μs, about

900μs, or about 1000μs, and a pulse frequency on the order of about 1-10 Hz. The polarity of the pulses may be reversed during the electroporation procedure by switching the connectors to the pulse generator. Pulses are repeated multiple times. The electroporation parameters (*e.g.* voltage amplitude, duration of pulse, number of pulses, depth of electrode insertion and frequency) will vary based on target tissue type, number of electrodes used and distance of electrode spacing, as would be understood by one of ordinary skill in the art.

Immediately after completion of the pulse regimen, subjects receiving electroporation can be optionally treated with membrane stabilizing agents to prolong cell membrane permeability as a result of the electroporation. Examples of membrane stabilizing agents include, but are not limited to, steroids (e.g. dexamethasone, methylprednisone and progesterone), angiotensin II and vitamin E. A single dose of dexamethasone, approximately 0.1 mg per kilogram of body weight, should be sufficient to achieve a beneficial affect.

EAPD techniques such as electroporation can also be used for plasmids contained in liposome formulations. The liposome – plasmid suspension is administered to the animal or patient and the site of injection is treated with a safe but effective electrical field generated, for example, by a TriGrid needle array. The electroporation may aid in plasmid delivery to the cell by destabilizing the liposome bilayer so that membrane fusion between the liposome and the target cellular structure occurs. Electroporation may also aid in plasmid delivery to the cell by triggering the release of the plasmid, in high concentrations, from the liposome at the surface of the target cell so that the plasmid is driven across the cell membrane by a concentration gradient via the pores created in the cell membrane as a result of the electroporation.

Female BALB/c mice aged 8-10 weeks are anesthetized with inhalant isoflurane and maintained under anesthesia for the duration of the electroporation procedure. The legs are shaved prior to treatment. Plasmid constructs comprising codon-optimized and non-codon-optimized coding

regions which encode LF, PA or various fragments, variants or derivatives, as described herein, are administered to BALB/c mice (n = 10) via unilateral injection in the quadriceps, with 50 μg total of a plasmid DNA per mouse, using an 0.3 cc insulin syringe and a 26 gauge, 1/2 length needle fitted with a plastic collar to regulate injection depth. Approximately one minute after injection, electrodes are applied. Modified caliper electrodes are used to apply the electrical pulse. *See* Hartikka J. *et al. Mol Ther 188*:407-415 (2001). The caliper electrode plates are coated with conductivity gel and applied to the sides of the injected muscle before closing to a gap of 3 mm for administration of pulses. EAPD is applied using a square pulse type at 1-10 Hz with a field strength of 100-500 V/cm, 1-10 pulses, of 10-100 ms each.

Mice are vaccinated ± EAPD at 0, 2 and 4 weeks. As endpoints, serum IgG titers against the various LF and PA antigens are measured as described in Example 9, as well as Letx neutralization titers.

Rabbits (n = 3) are given bilateral injections in the quadriceps muscle with plasmid constructs comprising codon-optimized and non-codon-optimized coding regions which encode LF, PA or various fragments, variants or derivatives, as described herein. The implantation area is shaved and the TriGrid electrode array is implanted into the target region of the muscle. 3.0 mg of plasmid DNA is administered per dose through the injection port of the electrode array. An injection collet is used to control the depth of injection. Electroporation begins approximately one minute after injection of the plasmid DNA is complete. Electroporation is administered with a TriGrid needle array, with electrodes evenly spaced 7mm apart, using an Ichor TGP-2 pulse generator. The array is inserted into the target muscle to a depth of about 1 to 2 cm. 4-8 pulses are administered. Each pulse has a duration of about 50-100 μs, an amplitude of about 1-1.2kV/cm and a pulse frequency of 1 Hz. The injection and electroporation may be repeated.

Sera are collected from vaccinated rabbits at various time point. As endpoints, serum IgG titers against the various LF and PA antigens are measured as described in Example 9, as well as Letx neutralization titers.

To test the effect of electroporation on therapeutic protein expression in non-human primates, male or female cynomonologous macques are given either 2 or 6 i.m. injections of plasmid constructs comprising codon-optimized and non-codon-optimized coding regions which encode LF, PA or various fragments, variants or derivatives, as described herein, (0.1 to 10 mg DNA total per animal). Target muscle groups include, but are not limited to, bilateral rectus fermoris, cranial tibialis, biceps, gastrocenemius or deltoid muscles. The target area is shaved and a needle array, comprising between 4 and 10 electrodes, spaced between 0.5-1.5 cm apart, is implanted into the target muscle. Once injections are complete, a sequence of brief electrical pulses are applied to the electrodes implanted in the target muscle using an Ichor TGP-2 pulse generator. The pulses have an amplitude of approximately 120 – 200V. The pulse sequence is completed within one second. During this time, the target muscle may make brief contractions or twitches. The injection and electroporation may be repeated.

Sera are collected from vaccinated monkeys at various time points. As endpoints, serum IgG titers against the various LF and PA antigens are measured as described in Example 9, as well as Letx neutralization titers.

***

The present invention is not to be limited in scope by the specific embodiments described which are intended as single illustrations of individual aspects of the invention, and any compositions or methods which are functionally equivalent are within the scope of this invention. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and accompanying drawings. Such modifications are intended to fall within the scope of the appended claims.

All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual

publication or patent application was specifically and individually indicated to be incorporated by reference.

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a nucleic acid fragment which encodes at least 50 contiguous amino acids of SEQ ID NO:4, wherein said nucleic acid fragment is a fragment of a codon-optimized coding region for the polypeptide of SEQ ID NO:4;

wherein about 11 of the 24 phenylalanine codons in said coding region are TTT and about 13 of said phenylalanine codons are TTC;

wherein about 5 of the 62 leucine codons in said coding region are TTA, about 8 of said leucine codons are TTG, about 8 of said leucine codons are CTT, about 12 of said leucine codons are CTC, about 4 of said leucine codons are CTA, and about 25 of said leucine codons are CTG;

wherein about 20 of the 57 isoleucine codons in said coding region are ATT, about 28 of said isoleucine codons are ATC, and about 9 of said isoleucine codons are ATA;

wherein the 10 methionine codons in said coding region are ATG;

wherein about 8 of the 43 valine codons in said coding region are GTT, about 10 of said valine codons are GTG, about 5 of said valine codons are GTA, and about 20 of said valine codons are GTG;

wherein about 13 of the 72 serine codons in said coding region are TCT, about 16 of said serine codons are TCC, about 11 of said serine codons are TCA, about 4 of said serine codons are TCG, about 11 of said serine codons are AGT, and about 17 of said serine codons are AGC;

wherein about 8 of the 29 proline codons in said coding region are CCT, about 10 of said proline codons are CCC, about 8 of said proline codons are CCA, and about 3 of said proline codons are CCG;

wherein about 14 of the 58 threonine codons in said coding region are ACT, about 21 of said threonine codons are ACC, about 16 of said threonine codons are ACA, and about 7 of said threonine codons are ACG;

wherein about 11 of the 41 alanine codons in said coding region are GGT, about 17 of said alanine codons are GCC, about 9 of said alanine codons are GCA, and about 4 of said alanine codons are GCG;

wherein about 12 of the 28 tyrosine codons in said coding region are TAT and about 16 of said tyrosine codons are TAC;

wherein about 4 of the 10 histidine codons in said coding region are CAT and about 6 of said histidine codons are CAC;

wherein about 8 of the 31 glutamine codons in said coding region are CAA and about 23 of said glutamine codons are CAG;

wherein about 32 of the 69 asparagine codons in said coding region are AAT and about 37 of said asparagine codons are AAC;

wherein about 25 of the 60 lysine codons in said coding region are AAA and about 35 of said lysine codons are AAG;

wherein about 22 of the 47 aspartic acid codons in said coding region are GAT and about 25 of said aspartic acid codons are GAC;

wherein about 21 of the 51 glutamic acid codons in said coding region are GAA and about 30 of said glutamic acid codons are GAG;

wherein the 7 tryptophan codons in said coding region are TGG;

wherein about 2 of the 29 arginine codons in said coding region are CGT, about 6 of said arginine codons are CGC, about 3 of said arginine codons are CGA, about 6 of said arginine codons are CGG, about 6 of said arginine codons are AGA, and about 6 of said arginine codons are AGG; and

wherein about 6 of the 36 glycine codons in said coding region are GGT, about 12 of said glycine codons are GGC, about 9 of said glycine codons are GGA, and about 9 of said glycine codons are GGG.

2. The polynucleotide of claim 1, wherein said nucleic acid fragment encodes at least 100 contiguous amino acids of SEQ ID NO:4.

3. The polynucleotide of claim 2, wherein said nucleic acid fragment encodes amino acids 199 to 764 of SEQ ID NO:4.

4. The polynucleotide of claim 3, wherein said nucleic acid fragment comprises nucleotides 82 to 1782 of SEQ ID NO:1.

5. The polynucleotide of any one of claims 1-4, wherein said nucleic acid fragment is ligated to a heterologous nucleic acid.

6. The polynucleotide of claim 5, wherein said heterologous nucleic acid encodes a heterologous polypeptide fused to the polypeptide encoded by said nucleic acid fragment.

7. The polynucleotide of claim 6, wherein said heterologous polypeptide is a secretory signal peptide.

8. The polynucleotide of claim 7, wherein said signal peptide is a human tissue plasminogen activator (hTPA) signal peptide.

9. The polynucleotide of claim 8, comprising nucleotides 13-1782 of SEQ ID NO:1.

10. The polynucleotide of claim 9, comprising SEQ ID NO:1.

11. The polynucleotide of claim 1, wherein said nucleic acid fragment encodes amino acids 30 to 764 of SEQ ID NO:4.

12. The polynucleotide of claim 11, wherein said nucleic acid fragment is ligated to a heterologous nucleic acid.

13. The polynucleotide of claim 12, wherein said heterologous nucleic acid encodes a heterologous polypeptide fused to the polypeptide encoded by said nucleic acid fragment.

14. The polynucleotide of claim 13, wherein said heterologous polypeptide is a secretory signal peptide.

15. The polynucleotide of claim 14, wherein said signal peptide is a human tissue plasminogen activator (hTPA) signal peptide.

16. The polynucleotide of claim 11, wherein said nucleic acid fragment comprises nucleotides 88 to 2292 of SEQ ID NO:23.

17. The polynucleotide of any one of claims 1-16, which is DNA, and wherein said nucleic acid fragment is operably associated with a promoter.

18. The polynucleotide of any one of claims 1-16, which is RNA.

19. The polynucleotide of claim 18, which is messenger RNA (mRNA).

20. A vector comprising the polynucleotide of any one of claims 1-17.

21. The vector of claim 20, which is a plasmid.

22. A composition comprising the polynucleotide of any one of claims 1-19, and a carrier.

23. The composition of claim 21, further comprising a component selected from the group consisting of an adjuvant, and a transfection facilitating compound.

24. The composition of claim 23, wherein said adjuvant is selected from the group consisting of:

(±)-N-(3-aminopropyl)-N,N-dimethyl-2,3-bis(*syn*-9-tetradeceneyloxy)-1-propanaminium bromide (GAP-DMORIE) and a neutral lipid;

a cytokine;

mono-phosphoryl lipid A and trehalosedicorynomycolateAF (MPL + TDM);

a solubilized mono-phosphoryl lipid A formulation; and

CRL1005/BAK.

25. The composition of claim 24, wherein said adjuvant comprises(±)-N-(3-aminopropyl)-N,N-dimethyl-2,3-bis(*syn*-9-

tetradeceneyloxy)-1-propanaminium bromide (GAP-DMORIE), and wherein said neutral lipid is selected from the group consisting of:

1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine (DOPE),

1,2-diphytanoyl-*sn*-glycero-3-phosphoethanolamine (DPyPE), and

1,2-dimyristoyl-glycer-3-phosphoethanolamine (DMPE).

26. The composition of claim 25, wherein said neutral lipid is DPyPE.

27. The composition of claim 23, comprising the transfection facilitating compound (±)-N-(2-hydroxyethyl)-N,N-dimethyl-2,3-bis(tetradecyloxy)-1-propanaminium bromide) (DMRIE).

28. A method to treat or prevent anthrax infection in a vertebrate comprising: administering to a vertebrate in need thereof the composition of any one of claims 22-27.

29. An isolated polynucleotide comprising a nucleic acid fragment which encodes a polypeptide at least 90% identical to amino acids 199 to 764 of SEQ ID NO:4, wherein said nucleic acid fragment is a variant fragment of an optimized coding region for the polypeptide of SEQ ID NO:4;

wherein about 11 of the 24 phenylalanine codons in said coding region are TTT and about 13 of said phenylalanine codons are TTC;

wherein about 5 of the 62 leucine codons in said coding region are TTA, about 8 of said leucine codons are TTG, about 8 of said leucine codons are CTT, about 12 of said leucine codons are CTC, about 4 of said leucine codons are CTA, and about 25 of said leucine codons are CTG;

wherein about 20 of the 57 isoleucine codons in said coding region are ATT, about 28 of said isoleucine codons are ATC, and about 9 of said isoleucine codons are ATA;

wherein the 10 methionine codons in said coding region are ATG;

wherein about 8 of the 43 valine codons in said coding region are GTT, about 10 of said valine codons are GTG, about 5 of said valine codons are GTA, and about 20 of said valine codons are GTG;

wherein about 13 of the 72 serine codons in said coding region are TCT, about 16 of said serine codons are TCC, about 11 of said serine codons are TCA, about 4 of said serine codons are TCG, about 11 of said serine codons are AGT, and about 17 of said serine codons are AGC;

wherein about 8 of the 29 proline codons in said coding region are CCT, about 10 of said proline codons are CCC, about 8 of said proline codons are CCA, and about 3 of said proline codons are CCG;

wherein about 14 of the 58 threonine codons in said coding region are ACT, about 21 of said threonine codons are ACC, about 16 of said threonine codons are ACA, and about 7 of said threonine codons are ACG;

wherein about 11 of the 41 alanine codons in said coding region are GGT, about 17 of said alanine codons are GCC, about 9 of said alanine codons are GCA, and about 4 of said alanine codons are GCG;

wherein about 12 of the 28 tyrosine codons in said coding region are TAT and about 16 of said tyrosine codons are TAC;

wherein about 4 of the 10 histidine codons in said coding region are CAT and about 6 of said histidine codons are CAC;

wherein about 8 of the 31 glutamine codons in said coding region are CAA and about 23 of said glutamine codons are CAG;

wherein about 32 of the 69 asparagine codons in said coding region are AAT and about 37 of said asparagine codons are AAC;

wherein about 25 of the 60 lysine codons in said coding region are AAA and about 35 of said lysine codons are AAG;

wherein about 22 of the 47 aspartic acid codons in said coding region are GAT and about 25 of said aspartic acid codons are GAC;

wherein about 21 of the 51 glutamic acid codons in said coding region are GAA and about 30 of said glutamic acid codons are GAG;

wherein the 7 tryptophan codons in said coding region are TGG;

wherein about 2 of the 29 arginine codons in said coding region are CGT, about 6 of said arginine codons are CGC, about 3 of said arginine codons are CGA, about 6 of said arginine codons are CGG, about 6 of said arginine codons are AGA, and about 6 of said arginine codons are AGG; and

wherein about 6 of the 36 glycine codons in said coding region are GGT, about 12 of said glycine codons are GGC, about 9 of said glycine codons are GGA, and about 9 of said glycine codons are GGG.

30. The polynucleotide of claim 29, wherein said nucleic acid fragment encodes a polypeptide at least 95% identical to amino acids 199 to 764 of SEQ ID NO:4.

31. The polynucleotide of claim 29, wherein the codons in said nucleic acid fragment corresponding to amino acids 342 and 343 of SEQ ID NO:4 are deleted.

32. The polynucleotide of claim 31, wherein said nucleic acid fragment encodes amino acids 24 to 587 of SEQ ID NO:6.

33. The polynucleotide of claim 32, which comprises nucleotides 82 to 1773 of SEQ ID NO:5.

34. The polynucleotide of claim 29, wherein the asparagine codons in said nucleic acid fragment corresponding to amino acids 275, 321, 357, 417, 505, 538, 599, 650, 693, and 738 of SEQ ID NO:4 are deleted and each of said asparagine codons is replaced with a codon which codes for an amino acid other than asparagine.

35. The polynucleotide of claim 34, wherein the asparagine codons in said nucleic acid fragment corresponding to amino acids 275, 321, 357, 417, 505, 538, 599, 650, 693, and 738 of SEQ ID NO:4 are deleted and each of said asparagine codons replaced with a codon which codes for glutamine.

36. The polynucleotide of claim 35, wherein said nucleic acid fragment encodes amino acids 24 to 589 of SEQ ID NO:18.

37. The polynucleotide of claim 36, which comprises nucleotides 82 to 1779 of SEQ ID NO:17.

38. The polynucleotide of claim 31, wherein the asparagine codons in said nucleic acid fragment corresponding to amino acids 275, 321, 357, 417, 505, 538, 599, 650, 693, and 738 of SEQ ID NO:4 are deleted and each of said asparagine codons is replaced with a codon which codes for an amino acid other than asparagine.

39. The polynucleotide of claim 38, wherein the asparagine codons in said nucleic acid fragment corresponding to amino acids 39, 153, 275, 321, 357, 417, 505, 538, 599, 650, 693, and 738 of SEQ ID NO:4 are deleted and each of said asparagine codons replaced with a codon which codes for glutamine.

40. The polynucleotide of any one of claims 29-39, wherein said nucleic acid fragment is ligated to a heterologous nucleic acid.

41. The polynucleotide of claim 40, wherein said heterologous nucleic acid encodes a heterologous polypeptide fused to the polypeptide encoded by said nucleic acid fragment.

42. The polynucleotide of claim 41, wherein said heterologous polypeptide is a secretory signal peptide.

43. The polynucleotide of claim 42, wherein said signal peptide is a human tissue plasminogen activator (hTPA) signal peptide.

44. The polynucleotide of any one of claims 29-43, which is DNA, and wherein said nucleic acid fragment is operably associated with a promoter.

45. The polynucleotide of any one of claims 29-43, which is RNA.

46. The polynucleotide of claim 45, which is messenger RNA (mRNA).

47. A vector comprising the polynucleotide of any one of claims 29-44.

48. The vector of claim 47, which is a plasmid.

49. A composition comprising the polynucleotide of any one of claims 29-46, and a carrier.

50. The composition of claim 49, further comprising a component selected from the group consisting of an adjuvant, and a transfection facilitating compound.

51. The composition of claim 50, wherein said adjuvant is selected from the group consisting of:

(±)-N-(3-aminopropyl)-N,N-dimethyl-2,3-bis(*syn*-9-tetradeceneyloxy)-1-propanaminium bromide (GAP-DMORIE) and a neutral lipid;

a cytokine;

mono-phosphoryl lipid A and trehalosedicorynomycolateAF (MPL + TDM);

a solubilized mono-phosphoryl lipid A formulation; and

CRL1005/BAK.

52. The composition of claim 51, wherein said adjuvant comprises(±)-N-(3-aminopropyl)-N,N-dimethyl-2,3-bis(*syn*-9-tetradeceneyloxy)-1-propanaminium bromide (GAP-DMORIE), and wherein said neutral lipid is selected from the group consisting of:

1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine (DOPE),

1,2-diphytanoyl-*sn*-glycero-3-phosphoethanolamine (DPyPE), and

1,2-dimyristoyl-glycer-3-phosphoethanolamine (DMPE).

53. The composition of claim 52, wherein said neutral lipid is DPyPE.

54. The composition of claim 50, comprising the transfection facilitating compound (±)-N-(2-hydroxyethyl)-N,N-dimethyl-2,3-bis(tetradecyloxy)-1-propanaminium bromide) (DMRIE).

55. A method to treat or prevent anthrax infection in a vertebrate comprising: administering to a vertebrate in need thereof the composition of any one of claims 49-54.

56. An isolated polynucleotide comprising a nucleic acid fragment which encodes a polypeptide at least 90% identical to amino acids 30 to 764 of SEQ ID NO:4, wherein said nucleic acid fragment is a variant fragment of an optimized coding region for the polypeptide of SEQ ID NO:4;

wherein about 11 of the 24 phenylalanine codons in said coding region are TTT and about 13 of said phenylalanine codons are TTC;

wherein about 5 of the 62 leucine codons in said coding region are TTA, about 8 of said leucine codons are TTG, about 8 of said leucine codons are CTT, about 12 of said leucine codons are CTC, about 4 of said leucine codons are CTA, and about 25 of said leucine codons are CTG;

wherein about 20 of the 57 isoleucine codons in said coding region are ATT, about 28 of said isoleucine codons are ATC, and about 9 of said isoleucine codons are ATA;

wherein the 10 methionine codons in said coding region are ATG;

wherein about 8 of the 43 valine codons in said coding region are GTT, about 10 of said valine codons are GTG, about 5 of said valine codons are GTA, and about 20 of said valine codons are GTG;

wherein about 13 of the 72 serine codons in said coding region are TCT, about 16 of said serine codons are TCC, about 11 of said serine codons are TCA, about 4 of said serine codons are TCG, about 11 of said serine codons are AGT, and about 17 of said serine codons are AGC;

wherein about 8 of the 29 proline codons in said coding region are CCT, about 10 of said proline codons are CCC, about 8 of said proline codons are CCA, and about 3 of said proline codons are CCG;

wherein about 14 of the 58 threonine codons in said coding region are ACT, about 21 of said threonine codons are ACC, about 16 of said threonine codons are ACA, and about 7 of said threonine codons are ACG;

wherein about 11 of the 41 alanine codons in said coding region are GGT, about 17 of said alanine codons are GCC, about 9 of said alanine codons are GCA, and about 4 of said alanine codons are GCG;

wherein about 12 of the 28 tyrosine codons in said coding region are TAT and about 16 of said tyrosine codons are TAC;

wherein about 4 of the 10 histidine codons in said coding region are CAT and about 6 of said histidine codons are CAC;

wherein about 8 of the 31 glutamine codons in said coding region are CAA and about 23 of said glutamine codons are CAG;

wherein about 32 of the 69 asparagine codons in said coding region are AAT and about 37 of said asparagine codons are AAC;

wherein about 25 of the 60 lysine codons in said coding region are AAA and about 35 of said lysine codons are AAG;

wherein about 22 of the 47 aspartic acid codons in said coding region are GAT and about 25 of said aspartic acid codons are GAC;

wherein about 21 of the 51 glutamic acid codons in said coding region are GAA and about 30 of said glutamic acid codons are GAG;

wherein the 7 tryptophan codons in said coding region are TGG;

wherein about 2 of the 29 arginine codons in said coding region are CGT, about 6 of said arginine codons are CGC, about 3 of said arginine codons are CGA, about 6 of said arginine codons are CGG, about 6 of said arginine codons are AGA, and about 6 of said arginine codons are AGG; and

wherein about 6 of the 36 glycine codons in said coding region are GGT, about 12 of said glycine codons are GGC, about 9 of said glycine codons are GGA, and about 9 of said glycine codons are GGG.

57. The polynucleotide of claim 56, wherein said nucleic acid fragment encodes a polypeptide at least 95% identical to amino acids 30 to 764 of SEQ ID NO:4.

58. The polynucleotide of claim 56, wherein the codons in said nucleic acid fragment corresponding to amino acids 192 to 197 of SEQ ID NO:4 are deleted.

59. The polynucleotide of claim 58, wherein said nucleic acid fragment encodes amino acids 24 to 752 of SEQ ID NO:8.

60. The polynucleotide of claim 59, which comprises nucleotides 82 to 2268 of SEQ ID NO:7.

61. The polynucleotide of claim 56, wherein the asparagine codons in said nucleic acid fragment corresponding to amino acids 39, 153, 275, 321, 357, 417, 505, 538, 599, 650, 693, and 738 of SEQ ID NO:4 are deleted and each of said asparagine codons is replaced with a codon which codes for an amino acid other than asparagine.

62. The polynucleotide of claim 61, wherein the asparagine codons in said nucleic acid fragment corresponding to amino acids 39, 153, 275, 321, 357, 417, 505, 538, 599, 650, 693, and 738 of SEQ ID NO:4 are deleted and each of said asparagine codons replaced with a codon which codes for glutamine.

63. The polynucleotide of claim 58, wherein the asparagine codons in said nucleic acid fragment corresponding to amino acids 39, 153, 275, 321, 357, 417, 505, 538, 599, 650, 693, and 738 of SEQ ID NO:4 are deleted and each of said asparagine codons is replaced with a codon which codes for an amino acid other than asparagine.

64. The polynucleotide of claim 63, wherein the asparagine codons in said nucleic acid fragment corresponding to amino acids 39, 153, 275, 321, 357, 417, 505, 538, 599, 650, 693, and 738 of SEQ ID NO:AP are deleted and each of said asparagine codons replaced with a codon which codes for glutamine.

65. The polynucleotide of any one of claims 56-64, wherein said nucleic acid fragment is ligated to a heterologous nucleic acid.

66. The polynucleotide of claim 65, wherein said heterologous nucleic acid encodes a heterologous polypeptide fused to the polypeptide encoded by said nucleic acid fragment.

67. The polynucleotide of claim 66, wherein said heterologous polypeptide is a secretory signal peptide.

68. The polynucleotide of claim 67, wherein said signal peptide is a human tissue plasminogen activator (hTPA) signal peptide.

69. The polynucleotide of any one of claims 56-68, which is DNA, and wherein said nucleic acid fragment is operably associated with a promoter.

70. The polynucleotide of any one of claims 56-68, which is RNA.

71. The polynucleotide of claim 70, which is messenger RNA (mRNA).

72. A vector comprising the polynucleotide of any one of claims 56-69.

73. The vector of claim 72, which is a plasmid.

74. A composition comprising the polynucleotide of any one of claims 56-71, and a carrier.

75. The composition of claim 74, further comprising a component selected from the group consisting of an adjuvant, and a transfection facilitating compound.

76. The composition of claim 75, wherein said adjuvant is selected from the group consisting of:

(±)-N-(3-aminopropyl)-N,N-dimethyl-2,3-bis(*syn*-9-tetradeceneyloxy)-1-propanaminium bromide (GAP-DMORIE) and a neutral lipid;

a cytokine;

mono-phosphoryl lipid A and trehalosedicorynomycolateAF (MPL + TDM);

a solubilized mono-phosphoryl lipid A formulation; and

CRL1005/BAK.

77. The composition of claim 76, wherein said adjuvant comprises(±)-N-(3-aminopropyl)-N,N-dimethyl-2,3-bis(*syn*-9-tetradeceneyloxy)-1-propanaminium bromide (GAP-DMORIE), and wherein said neutral lipid is selected from the group consisting of:

1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine (DOPE),

1,2-diphytanoyl-*sn*-glycero-3-phosphoethanolamine (DPyPE), and

1,2-dimyristoyl-glycer-3-phosphoethanolamine (DMPE).

78. The composition of claim 77, wherein said neutral lipid is DPyPE.

79. The composition of claim 75, comprising the transfection facilitating compound (±)-N-(2-hydroxyethyl)-N,N-dimethyl-2,3-bis(tetradecyloxy)-1-propanaminium bromide) (DMRIE).

80. A method to treat or prevent anthrax infection in a vertebrate comprising: administering to a vertebrate in need thereof the composition of any one of claims 74-79.

81. An isolated polynucleotide comprising a nucleic acid fragment which encodes at least 50 contiguous amino acids of SEQ ID NO:12, wherein said nucleic acid fragment is a portion of an optimized coding region for the polypeptide of SEQ ID NO:12;

wherein about 13 of the 29 phenylalanine codons in said coding region are TTT and about 16 of said phenylalanine codons are TTC;

wherein about 6 of the 80 leucine codons in said coding region are TTA, about 10 of said leucine codons are TTG, about 10 of said leucine codons are CTT, about 16 of said leucine codons are CTC, about 6 of said leucine codons are CTA, and about 32 of said leucine codons are CTG;

wherein about 26 of the 74 isoleucine codons in said coding region are ATT, about 36 of said isoleucine codons are ATC, and about 12 of said isoleucine codons are ATA;

wherein the 10 methionine codons in said coding region are ATG;

wherein about 7 of the 40 valine codons in said coding region are GTT, about 9 of said valine codons are GTG, about 5 of said valine codons are GTA, and about 19 of said valine codons are GTG;

wherein about 10 of the 54 serine codons in said coding region are TCT, about 12 of said serine codons are TCC, about 8 of said serine codons are TCA, about 3 of said serine codons are TCG, about 8 of said serine codons are AGT, and about 13 of said serine codons are AGC;

wherein about 6 of the 21 proline codons in said coding region are CCT, about 7 of said proline codons are CCC, about 6 of said proline codons are CCA, and about 2 of said proline codons are CCG;

wherein about 7 of the 28 threonine codons in said coding region are ACT, about 10 of said threonine codons are ACC, about 8 of said threonine codons are ACA, and about 3 of said threonine codons are ACG;

wherein about 9 of the 34 alanine codons in said coding region are GGT, about 14 of said alanine codons are GCC, about 8 of said alanine codons are GCA, and about 3 of said alanine codons are GCG;

wherein about 15 of the 35 tyrosine codons in said coding region are TAT and about 20 of said tyrosine codons are TAC;

wherein about 9 of the 21 histidine codons in said coding region are CAT and about 12 of said histidine codons are CAC;

wherein about 10 of the 41 glutamine codons in said coding region are CAA and about 31 of said glutamine codons are CAG;

wherein about 25 of the 54 asparagine codons in said coding region are AAT and about 29 of said asparagine codons are AAC;

wherein about 36 of the 86 lysine codons in said coding region are AAA and about 50 of said lysine codons are AAG;

wherein about 25 of the 55 aspartic acid codons in said coding region are GAT and about 30 of said aspartic acid codons are GAC;

wherein about 33 of the 79 glutamic acid codons in said coding region are GAA and about 46 of said glutamic acid codons are GAG;

wherein the single cysteine codon in said coding region is selected from the group consisting of TGT and TGC;

wherein the 5 tryptophan codons in said coding region are TGG;

wherein about 2 of the 27 arginine codons in said coding region are CGT, about 5 of said arginine codons are CGC, about 3 of said arginine codons are CGA, about 6 of said arginine codons are CGG, about 6 of said arginine codons are AGA, and about 5 of said arginine codons are AGG; and

wherein about 6 of the 35 glycine codons in said coding region are GGT, about 12 of said glycine codons are GGC, about 8 of said glycine codons are GGA, and about 9 of said glycine codons are GGG.

82. The polynucleotide of claim 81, wherein said cysteine codon in said coding region is TGT.

83. The polynucleotide of claim 81, wherein said cysteine codon in said coding region is TGC.

84. The polynucleotide of claim 81, wherein said nucleic acid fragment encodes at least 100 contiguous amino acids of SEQ ID NO:12.

85. The polynucleotide of claim 84, wherein said nucleic acid fragment encodes amino acids 34 to 809 of SEQ ID NO:12.

86. The polynucleotide of claim 85, wherein said nucleic acid fragment comprises nucleotides 99 to 2427 of SEQ ID NO:26.

87. The polynucleotide of claim 81, wherein said nucleic acid fragment encodes amino acids 34-583 of SEQ ID NO:12.

88. The polynucleotide of claim 81, wherein said nucleic acid fragment encodes amino acids 34-254 of SEQ ID NO:12.

89. The polynucleotide of claim 81, wherein said nucleic acid fragment encodes amino acids 34-295 of SEQ ID NO:12.

90. The polynucleotide of any one of claims 81-89, wherein said nucleic acid fragment is ligated to a heterologous nucleic acid.

91. The polynucleotide of claim 90, wherein said heterologous nucleic acid encodes a heterologous polypeptide fused to the polypeptide encoded by said nucleic acid fragment.

92. The polynucleotide of claim 91, wherein said heterologous polypeptide is a secretory signal peptide.

93. The polynucleotide of claim 92, wherein said signal peptide is a human tissue plasminogen activator (hTPA) signal peptide.

94. The polynucleotide of any one of claims 81-93, which is DNA, and wherein said nucleic acid fragment is operably associated with a promoter.

95. The polynucleotide of any one of claims 81-93, which is RNA.

96. The polynucleotide of claim 95, which is messenger RNA (mRNA).

97. A vector comprising the polynucleotide of any one of claims 81-94.

98. The vector of claim 97, which is a plasmid.

99. A composition comprising the polynucleotide of any one of claims 81-96, and a carrier.

100. The composition of claim 99, further comprising a component selected from the group consisting of an adjuvant, and a transfection facilitating compound.

101. The composition of claim 100, wherein said adjuvant is selected from the group consisting of:

(±)-N-(3-aminopropyl)-N,N-dimethyl-2,3-bis(*syn*-9-tetradeceneyloxy)-1-propanaminium bromide (GAP-DMORIE) and a neutral lipid;

a cytokine;

mono-phosphoryl lipid A and trehalosedicorynomycolateAF (MPL + TDM);

a solubilized mono-phosphoryl lipid A formulation; and

CRL1005/BAK.

102. The composition of claim 101, wherein said adjuvant comprises(±)-N-(3-aminopropyl)-N,N-dimethyl-2,3-bis(*syn*-9-tetradeceneyloxy)-1-propanaminium bromide (GAP-DMORIE), and wherein said neutral lipid is selected from the group consisting of:

1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine (DOPE),

1,2-diphytanoyl-*sn*-glycero-3-phosphoethanolamine (DPyPE), and

1,2-dimyristoyl-glycer-3-phosphoethanolamine (DMPE).

103. The composition of claim 102, wherein said neutral lipid is DPyPE.

104. The composition of claim 100, comprising the transfection facilitating compound (±)-N-(2-hydroxyethyl)-N,N-dimethyl-2,3-bis(tetradecyloxy)-1-propanaminium bromide) (DMRIE).

105. A method to treat or prevent anthrax infection in a vertebrate comprising: administering to a vertebrate in need thereof the composition of any one of claims 99-104.

106. An isolated polynucleotide comprising a nucleic acid fragment which encodes a polypeptide at least 90% identical to amino acids 34 to 809 of SEQ ID NO:12, wherein said nucleic acid fragment is a variant fragment of an optimized coding region for the polypeptide of SEQ ID NO:12;

wherein about 13 of the 29 phenylalanine codons in said coding region are TTT and about 16 of said phenylalanine codons are TTC;

wherein about 6 of the 80 leucine codons in said coding region are TTA, about 10 of said leucine codons are TTG, about 10 of said leucine codons are CTT, about 16 of said leucine codons are CTC, about 6 of said leucine codons are CTA, and about 32 of said leucine codons are CTG;

wherein about 26 of the 74 isoleucine codons in said coding region are ATT, about 36 of said isoleucine codons are ATC, and about 12 of said isoleucine codons are ATA;

wherein the 10 methionine codons in said coding region are ATG;

wherein about 7 of the 40 valine codons in said coding region are GTT, about 9 of said valine codons are GTG, about 5 of said valine codons are GTA, and about 19 of said valine codons are GTG;

wherein about 10 of the 54 serine codons in said coding region are TCT, about 12 of said serine codons are TCC, about 8 of said serine codons are TCA, about 3 of said serine codons are TCG, about 8 of said serine codons are AGT, and about 13 of said serine codons are AGC;

wherein about 6 of the 21 proline codons in said coding region are CCT, about 7 of said proline codons are CCC, about 6 of said proline codons are CCA, and about 2 of said proline codons are CCG;

wherein about 7 of the 28 threonine codons in said coding region are ACT, about 10 of said threonine codons are ACC, about 8 of said threonine codons are ACA, and about 3 of said threonine codons are ACG;

wherein about 9 of the 34 alanine codons in said coding region are GGT, about 14 of said alanine codons are GCC, about 8 of said alanine codons are GCA, and about 3 of said alanine codons are GCG;

wherein about 15 of the 35 tyrosine codons in said coding region are TAT and about 20 of said tyrosine codons are TAC;

wherein about 9 of the 21 histidine codons in said coding region are CAT and about 12 of said histidine codons are CAC;

wherein about 10 of the 41 glutamine codons in said coding region are CAA and about 31 of said glutamine codons are CAG;

wherein about 25 of the 54 asparagine codons in said coding region are AAT and about 29 of said asparagine codons are AAC;

wherein about 36 of the 86 lysine codons in said coding region are AAA and about 50 of said lysine codons are AAG;

wherein about 25 of the 55 aspartic acid codons in said coding region are GAT and about 30 of said aspartic acid codons are GAC;

wherein about 33 of the 79 glutamic acid codons in said coding region are GAA and about 46 of said glutamic acid codons are GAG;

wherein the single cysteine codon in said coding region is selected from the group consisting of TGT and TGC;

wherein the 5 tryptophan codons in said coding region are TGG;

wherein about 2 of the 27 arginine codons in said coding region are CGT, about 5 of said arginine codons are CGC, about 3 of said arginine

codons are CGA, about 6 of said arginine codons are CGG, about 6 of said arginine codons are AGA, and about 5 of said arginine codons are AGG; and

wherein about 6 of the 35 glycine codons in said coding region are GGT, about 12 of said glycine codons are GGC, about 8 of said glycine codons are GGA, and about 9 of said glycine codons are GGG.

107. The polynucleotide of claim 106, wherein said cysteine codon in said coding region is TGT.

108. The polynucleotide of claim 106, wherein said cysteine codon in said coding region is TGC.

109. The polynucleotide of claim 106, wherein nucleic acid fragment encodes a polypeptide at least 95% identical to amino acids 34 to 809 of SEQ ID NO:12.

110. The polynucleotide of claim 106, wherein the histidine codons in said nucleic acid fragment corresponding to positions 719 and 723 of SEQ ID NO:12 are deleted and each is replaced with a codon which codes for an amino acid other than histidine, and wherein the glutamic acid codon in said nucleic acid fragment corresponding to position 720 of SEQ ID NO:12 is deleted and replaced with a codon which codes for an amino acid other than glutamic acid.

111. The polynucleotide of claim 110, wherein the histidine codons in said nucleic acid fragment corresponding to positions 719 and 723 of SEQ ID NO:12 are deleted and each is replaced with an alanine codon selected from the group consisting of GCT, GCC, GCA, and GCG, and wherein the glutamic acid codon in said nucleic acid fragment corresponding to position 720 of SEQ ID NO:12 is deleted and replaced with an aspartic acid codon selected from the group consisting of GAT and GAC.

112. The polynucleotide of claim 111, which comprises nucleotides 82 to 2409 of SEQ ID NO:9.

113. The polynucleotide of claim 106, wherein the asparagine codons in said nucleic acid fragment corresponding to positions 62, 212, 286, 478, 712, 736, and 757 of SEQ ID NO:12 are deleted and each is replaced with a codon which codes for an amino acid other than asparagine.

114. The polynucleotide of claim 113, wherein the asparagine codons in said nucleic acid fragment corresponding to positions 62, 212, 286, 478, 712, 736, and 757 of SEQ ID NO:12 are deleted and each is replaced with a glutamine codon selected from the group consisting of CAA and CAG.

115. The polynucleotide of claim 110, wherein the asparagine codons in said nucleic acid fragment corresponding to positions 62, 212, 286, 478, 712, 736, and 757 of SEQ ID NO:12 are deleted and each is replaced with a codon which codes for an amino acid other than asparagine.

116. The polynucleotide of claim 115, wherein the asparagine codons in said nucleic acid fragment corresponding to positions 62, 212, 286, 478, 712, 736, and 757 of SEQ ID NO:12 are deleted and each is replaced with a glutamine codon selected from the group consisting of CAA and CAG.

117. The polynucleotide of claim 116, wherein said glutamine codon is CAA.

118. The polynucleotide of claim 116, wherein said glutamine codon is CAG.

119. The polynucleotide of claim 116, which comprises nucleotides 82 to 2409 of SEQ ID NO:19.

120. The polynucleotide of any one of claims 106-119, wherein said nucleic acid fragment is ligated to a heterologous nucleic acid.

121. The polynucleotide of claim 120, wherein said heterologous nucleic acid encodes a heterologous polypeptide fused to the polypeptide encoded by said nucleic acid fragment.

122. The polynucleotide of claim 121, wherein said heterologous polypeptide is a secretory signal peptide.

123. The polynucleotide of claim 122, wherein said signal peptide is a human tissue plasminogen activator (hTPA) signal peptide.

124. The polynucleotide of any one of claims 106-123, which is DNA, and wherein said nucleic acid fragment is operably associated with a promoter.

125. The polynucleotide of any one of claims 106-123, which is RNA.

126. The polynucleotide of claim 125, which is messenger RNA (mRNA).

127. A vector comprising the polynucleotide of any one of claims 106-124.

128. The vector of claim 127, which is a plasmid.

129. A composition comprising the polynucleotide of any one of claims 106-126, and a carrier.

130. The composition of claim 129, further comprising a component selected from the group consisting of an adjuvant, and a transfection facilitating compound.

131. The composition of claim 130, wherein said adjuvant is selected from the group consisting of:

(±)-N-(3-aminopropyl)-N,N-dimethyl-2,3-bis(*syn*-9-tetradeceneyloxy)-1-propanaminium bromide (GAP-DMORIE) and a neutral lipid;

a cytokine;

mono-phosphoryl lipid A and trehalosedicorynomycolateAF (MPL + TDM);

a solubilized mono-phosphoryl lipid A formulation; and

CRL1005/BAK.

132. The composition of claim 131, wherein said adjuvant comprises(±)-N-(3-aminopropyl)-N,N-dimethyl-2,3-bis(*syn*-9-tetradeceneyloxy)-1-propanaminium bromide (GAP-DMORIE), and wherein said neutral lipid is selected from the group consisting of:

1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine (DOPE), 1,2-diphytanoyl-*sn*-glycero-3-phosphoethanolamine (DPyPE), and 1,2-dimyristoyl-glycer-3-phosphoethanolamine (DMPE).

133. The composition of claim 132, wherein said neutral lipid is DPyPE.

134. The composition of claim 130, comprising the transfection facilitating compound (±)-N-(2-hydroxyethyl)-N,N-dimethyl-2,3-bis(tetradecyloxy)-1-propanaminium bromide) (DMRIE).

135. A method to treat or prevent anthrax infection in a vertebrate comprising: administering to a vertebrate in need thereof the composition of any one of claims 129-134.

Figure 1A

```
   1  gatatcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga
                   m  d   a  m  k  r   g  l  c   c  v  l   l  l  c  g

  61  gcagtcttcg tttcgcccag cagcgctggg ccaactgtgc ccgacagaga caatgatgga
        a  v  f   v  s  p   s  s  a  g   p  t  v   p  d  r   d  n  d  g

 121  atccctgata gtctagaggt tgagggatac acggtagatg tcaagaacaa aaggactttt
        i  p  d   s  l  e   v  e  g  y   t  v  d   v  k  n   k  r  t  f

 181  ctctcgcctt ggatctcaaa tatccatgag aagaaggggc ttaccaagta caagtcctcc
        l  s  p   w  i  s   n  i  h  e   k  k  g   l  t  k   y  k  s  s

 241  cccgagaagt ggtctaccgc ttccgatcca tatagcgatt tcgagaaggt cacaggccgg
        p  e  k   w  s  t   a  s  d  p   y  s  d   f  e  k   v  t  g  r

 301  atcgataaaa atgtgtctcc agaggctaga cacccccctgg tagcagccta cccgattgta
        i  d  k   n  v  s   p  e  a  r   h  p  l   v  a  a   y  p  i  v

 361  cacgtggaca tggagaacat cattctaagc aaaaacgagg accagtccac acaaaacact
        h  v  d   m  e  n   i  i  l  s   k  n  e   d  q  s   t  q  n  t

 421  gactccgaga cccgcaccat atctaaaaac accagtactt caaggaccca cacctctgaa
        d  s  e   t  r  t   i  s  k  n   t  s  t   s  r  t   h  t  s  e

 481  gtgcacggca atgcggaagt ccatgcatcg tttttcgata ttggtggctc cgtgtcagcc
        v  h  g   n  a  e   v  h  a  s   f  f  d   i  g  g   s  v  s  a

 541  ggctttagca atagcaactc ctcgacggtt gccattgacc actcactgtc attagcaggt
        g  f  s   n  s  n   s  s  t  v   a  i  d   h  s  l   s  l  a  g

 601  gagaggactt gggctgaaac tatgggtctg aataccgccg atacggcccg gctcaacgca
        e  r  t   w  a  e   t  m  g  l   n  t  a   d  t  a   r  l  n  a

 661  aatattcggt acgtcaacac agggactgct cctatatata acgtgctgcc tacgacaagt
        n  i  r   y  v  n   t  g  t  a   p  i  y   n  v  l   p  t  t  s

 721  cttgtcctgg gcaaaaatca gaccctcgca accattaagg caaaggaaaa tcagctgagc
        l  v  l   g  k  n   q  t  l  a   t  i  k   a  k  e   n  q  l  s

 781  cagatcctcg cccctaacaa ctattatcca tccaaaaatt tagcccccat agccctgaac
        q  i  l   a  p  n   n  y  y  p   s  k  n   l  a  p   i  a  l  n

 841  gcccaggacg acttttcctc tacccccata actatgaatt acaatcagtt cctggagctg
        a  q  d   d  f  s   s  t  p  i   t  m  n   y  n  q   f  l  e  l

 901  gaaaagacga agcagctgag actagacacc gatcaggtgt atggaaacat agcgacatat
        e  k  t   k  q  l   r  l  d  t   d  q  v   y  g  n   i  a  t  y

 961  aactttgaga acggccgcgt gcgcgtcgac actgggtcaa actggtctga agttctgccg
        n  f  e   n  g  r   v  r  v  d   t  g  s   n  w  s   e  v  l  p

1021  caaattcaag agacaaccgc cagaattatc tttaatggga aggacttgaa ccttgtcgaa
        q  i  q   e  t  t   a  r  i  i   f  n  g   k  d  l   n  l  v  e
```

Figure 1B

```
1081  cgtagaattg ccgccgtgaa ccccagtgat ccactcgaga cgactaaacc ggatatgaca
       r  r  i   a  a  v   n  p  s  d   p  l  e   t  t  k   p  d  m  t

1141  ctgaaagagg ctctgaagat tgccttcgga ttcaacgaac ctaatggcaa tttgcagtat
       l  k  e   a  l  k   i  a  f  g   f  n  e   p  n  g   n  l  q  y

1201  caggggaaag acatcacaga gtttgatttc aatttcgatc agcagacttc ccaaaatatc
       q  g  k   d  i  t   e  f  d  f   n  f  d   q  q  t   s  q  n  i

1261  aaaaatcagt tggcagagct gaatgccacc aatatctaca cggttctcga taaaatcaaa
       k  n  q   l  a  e   l  n  a  t   n  i  y   t  v  l   d  k  i  k

1321  cttaacgcca agatgaacat attgattcga gacaaacgct tccactacga ccgcaacaat
       l  n  a   k  m  n   i  l  i  r   d  k  r   f  h  y   d  r  n  n

1381  atagccgtag gcgctgatga gtctgtcgtc aaggaggctc atagggaagt tatcaacagc
       i  a  v   g  a  d   e  s  v  v   k  e  a   h  r  e   v  i  n  s

1441  agtactgaag ggctgttact taatatcgac aaggacattc ggaagatcct gtccgggtat
       s  t  e   g  l  l   l  n  i  d   k  d  i   r  k  i   l  s  g  y

1501  atcgtggaga tcgaggatac cgagggcctg aaggaagtca ttaacgaccg ctatgatatg
       i  v  e   i  e  d   t  e  g  l   k  e  v   i  n  d   r  y  d  m

1561  ctgaacattt ccagcttacg acaggacggt aagacattta ttgactttaa aaagtataac
       l  n  i   s  s  l   r  q  d  g   k  t  f   i  d  f   k  k  y  n

1621  gacaagctac ccctgtacat ttccaaccca aattacaaag ttaatgtgta tgctgtaacc
       d  k  l   p  l  y   i  s  n  p   n  y  k   v  n  v   y  a  v  t

1681  aaggagaaca caatcatcaa tccaagcgag aacggcgata ccagcacaaa tggaatcaaa
       k  e  n   t  i  i   n  p  s  e   n  g  d   t  s  t   n  g  i  k

1741  aagatcctta tatttagtaa aaaaggctac gagatcggtt gaggatcc
       k  i  l   i  f  s   k  k  g  y   e  i  g   -
```

Figure 2A

```
1    gatatcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga gcagtcttcg
                          m  d  a  m  k  r  g  l  c  c  v  l  l  l  c  g  a  v  f

71   tttcgcccag cagcgctggg ccaactgtgc ccgacagaga caatgatgga atccctgata gtctagaggt
      v  s  p  s  s  a  g  p  t  v  p  d  r  d  n  d  g  i  p  d  s  l  e

141  tgagggatac acggtagatg tcaagaacaa aaggactttt ctctcgcctt ggatctcaaa tatccatgag
     v  e  g  y  t  v  d  v  k  n  k  r  t  f  l  s  p  w  i  s  n  i  h  e

211  aagaaggggc ttaccaagta caagtcctcc cccgagaagt ggtctaccgc ttccgatcca tatagcgatt
      k  k  g  l  t  k  y  k  s  s  p  e  k  w  s  t  a  s  d  p  y  s  d

281  tcgagaaggt cacaggccgg atcgataaaa atgtgtctcc agaggctaga caccccctgg tagcagccta
      f  e  k  v  t  g  r  i  d  k  n  v  s  p  e  a  r  h  p  l  v  a  a

351  cccgattgta cacgtggaca tggagaacat cattctaagc aaaaacgagg accagtccac acaaaacact
     y  p  i  v  h  v  d  m  e  n  i  i  l  s  k  n  e  d  q  s  t  q  n  t

421  gactccgaga cccgcaccat atctaaaaac accagtactt caaggaccca cacctctgaa gtgcacggca
      d  s  e  t  r  t  i  s  k  n  t  s  t  s  r  t  h  t  s  e  v  h  g

491  atgcggaagt ccatgcatcg gatattggtg gctccgtgtc agccggcttt agcaatagca actcctcgac
      n  a  e  v  h  a  s  d  i  g  g  s  v  s  a  g  f  s  n  s  n  s  s

561  ggttgccatt gaccactcac tgtcattagc aggtgagagg acttgggctg aaactatggg tctgaatacc
     t  v  a  i  d  h  s  l  s  l  a  g  e  r  t  w  a  e  t  m  g  l  n  t

631  gccgatacgg cccggctcaa cgcaaatatt cggtacgtca acacagggac tgctcctata tataacgtgc
      a  d  t  a  r  l  n  a  n  i  r  y  v  n  t  g  t  a  p  i  y  n  v

701  tgcctacgac aagtcttgtc ctgggcaaaa atcagaccct cgcaaccatt aaggcaaagg aaaatcagct
      l  p  t  t  s  l  v  l  g  k  n  q  t  l  a  t  i  k  a  k  e  n  q

771  gagccagatc ctcgccccta acaactatta tccatccaaa aatttagccc ccatagccct gaacgcccag
     l  s  q  i  l  a  p  n  n  y  y  p  s  k  n  l  a  p  i  a  l  n  a  q

841  gacgactttt cctctacccc cataactatg aattacaatc agttcctgga gctggaaaag acgaagcagc
      d  d  f  s  s  t  p  i  t  m  n  y  n  q  f  l  e  l  e  k  t  k  q

911  tgagactaga caccgatcag gtgtatggaa acatagcgac atataacttt gagaacggcc gcgtgcgcgt
      l  r  l  d  t  d  q  v  y  g  n  i  a  t  y  n  f  e  n  g  r  v  r

981  cgacactggg tcaaactggt ctgaagttct gccgcaaatt caagagacaa ccgccagaat tatctttaat
     v  d  t  g  s  n  w  s  e  v  l  p  q  i  q  e  t  t  a  r  i  i  f  n

1051 gggaaggact tgaaccttgt cgaacgtaga attgccgccg tgaaccccag tgatccactc gagacgacta
      g  k  d  l  n  l  v  e  r  r  i  a  a  v  n  p  s  d  p  l  e  t  t

1121 aaccggatat gacactgaaa gaggctctga agattgcctt cggattcaac gaacctaatg gcaatttgca
      k  p  d  m  t  l  k  e  a  l  k  i  a  f  g  f  n  e  p  n  g  n  l

1191 gtatcagggg aaagacatca cagagtttga tttcaatttc gatcagcaga cttcccaaaa tatcaaaaat
     q  y  q  g  k  d  i  t  e  f  d  f  n  f  d  q  q  t  s  q  n  i  k  n

1261 cagttggcag agctgaatgc caccaatatc tacacggttc tcgataaaat caaacttaac gccaagatga
      q  l  a  e  l  n  a  t  n  i  y  t  v  l  d  k  i  k  l  n  a  k  m

1331 acatattgat tcgagacaaa cgcttccact acgaccgcaa caatatagcc gtaggcgctg atgagtctgt
     n  i  l  i  r  d  k  r  f  h  y  d  r  n  n  i  a  v  g  a  d  e  s

1401 cgtcaaggag gctcataggg aagttatcaa cagcagtact gaagggctgt tacttaatat cgacaaggac
     v  v  k  e  a  h  r  e  v  i  n  s  s  t  e  g  l  l  l  n  i  d  k  d

1471 attcggaaga tcctgtccgg gtatatcgtg gagatcgagg ataccgaggg cctgaaggaa gtcattaacg
      i  r  k  i  l  s  g  y  i  v  e  i  e  d  t  e  g  l  k  e  v  i  n
```

Figure 2B

```
1541  accgctatga tatgctgaac atttccagct tacgacagga cggtaagaca tttattgact ttaaaaagta
        d  r  y   d  m  l  n   i  s  s   l  r  q   d  g  k  t   f  i  d   f  k  k

1611  taacgacaag ctacccctgt acatttccaa cccaaattac aaagttaatg tgtatgctgt aaccaaggag
      y  n  d  k   l  p  l   y  i  s   n  p  n  y   k  v  n   v  y  a   v  t  k  e

1681  aacacaatca tcaatccaag cgagaacggc gataccagca caaatggaat caaaaagatc cttatattta
        n  t  i   i  n  p   s  e  n  g   d  t  s   t  n  g   i  k  k  i   l  i  f

1751  gtaaaaaagg ctacgagatc ggttgaggat cc
        s  k  k   g  y  e  i   g  -
```

Figure 3A

```
   1  gatatcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga
                         m  d   a  m  k  r   g  l  c   c  v  l   l  l  c  g

  61  gcagtcttcg tttcgcccag cgaagtgaag caagaaaatc gacttctgaa cgagagcgaa
        a  v  f   v  s  p   s  e  v  k   q  e  n   r  l  l   n  e  s  e

 121  agttcatcac agggtcttct cggatactac ttcagtgact tgaatttcca agcaccaatg
        s  s  s   q  g  l   l  g  y  y   f  s  d   l  n  f   q  a  p  m

 181  gtggtgacta gtagcaccac cggcgatttg agcattccca gctctgagtt ggagaacatt
        v  v  t   s  s  t   t  g  d  l   s  i  p   s  s  e   l  e  n  i

 241  cccagcgaaa atcagtactt ccagtctgct atctggtccg gattcattaa ggttaaaaag
        p  s  e   n  q  y   f  q  s  a   i  w  s   g  f  i   k  v  k  k

 301  tccgacgaat atacatttgc tacctcggcg gataaccatg tgacaatgtg ggtggacgac
        s  d  e   y  t  f   a  t  s  a   d  n  h   v  t  m   w  v  d  d

 361  caggaagtga tcaacaaggc ttcaaactct aataaaatcc ggctcgagaa ggggaggctc
        q  e  v   i  n  k   a  s  n  s   n  k  i   r  l  e   k  g  r  l

 421  taccagatca aaattcagta ccagcgggaa aaccctacag aaaaaggact cgatttcaag
        y  q  i   k  i  q   y  q  r  e   n  p  t   e  k  g   l  d  f  k

 481  ctgtactgga cagatagcca aaacaagaaa gaagttatca gctcagacaa tctgcagtta
        l  y  w   t  d  s   q  n  k  k   e  v  i   s  s  d   n  l  q  l

 541  cccgagctca agcagaagag ttctaataca agcgctgggc caactgtgcc cgacagagac
        p  e  l   k  q  k   s  s  n  t   s  a  g   p  t  v   p  d  r  d

 601  aatgatggaa tccctgatag tctagaggtt gagggataca cggtagatgt caagaacaaa
        n  d  g   i  p  d   s  l  e  v   e  g  y   t  v  d   v  k  n  k

 661  aggacttttc tctcgccttg gatctcaaat atccatgaga agaaggggct taccaagtac
        r  t  f   l  s  p   w  i  s  n   i  h  e   k  k  g   l  t  k  y

 721  aagtcctccc ccgagaagtg gtctaccgct tccgatccat atagcgattt cgagaaggtc
        k  s  s   p  e  k   w  s  t  a   s  d  p   y  s  d   f  e  k  v

 781  acaggccgga tcgataaaaa tgtgtctcca gaggctagac accccctggt agcagcctac
        t  g  r   i  d  k   n  v  s  p   e  a  r   h  p  l   v  a  a  y

 841  ccgattgtac acgtggacat ggagaacatc attctaagca aaaacgagga ccagtccaca
        p  i  v   h  v  d   m  e  n  i   i  l  s   k  n  e   d  q  s  t

 901  caaaacactg actccgagac ccgcaccata tctaaaaaca ccagtacttc aaggacccac
        q  n  t   d  s  e   t  r  t  i   s  k  n   t  s  t   s  r  t  h

 961  acctctgaag tgcacggcaa tgcggaagtc catgcatcgt tttcgatat tggtggctcc
        t  s  e   v  h  g   n  a  e  v   h  a  s   f  f  d   i  g  g  s

1021  gtgtcagccg gctttagcaa tagcaactcc tcgacggttg ccattgacca ctcactgtca
        v  s  a   g  f  s   n  s  n  s   s  t  v   a  i  d   h  s  l  s
```

Figure 3B

```
1081  ttagcaggtg agaggacttg ggctgaaact atgggtctga ataccgccga tacggcccgg
        l  a  g   e  r  t   w  a  e  t   m  g  l   n  t  a   d  t  a  r

1141  ctcaacgcaa atattcggta cgtcaacaca gggactgctc ctatatataa cgtgctgcct
        l  n  a   n  i  r   y  v  n  t   g  t  a   p  i  y   n  v  l  p

1201  acgacaagtc ttgtcctggg caaaaatcag accctcgcaa ccattaaggc aaaggaaaat
        t  t  s   l  v  l   g  k  n  q   t  l  a   t  i  k   a  k  e  n

1261  cagctgagcc agatcctcgc cctaacaac tattatccat ccaaaaattt agccccata
        q  l  s   q  i  l   a  p  n  n   y  y  p   s  k  n   l  a  p  i

1321  gccctgaacg cccaggacga cttttcctct acccccataa ctatgaatta caatcagttc
        a  l  n   a  q  d   d  f  s  s   t  p  i   t  m  n   y  n  q  f

1381  ctggagctgg aaaagacgaa gcagctgaga ctagacaccg atcaggtgta tggaaacata
        l  e  l   e  k  t   k  q  l  r   l  d  t   d  q  v   y  g  n  i

1441  gcgacatata actttgagaa cggccgcgtg cgcgtcgaca ctgggtcaaa ctggtctgaa
        a  t  y   n  f  e   n  g  r  v   r  v  d   t  g  s   n  w  s  e

1501  gttctgccgc aaattcaaga gacaaccgcc agaattatct ttaatgggaa ggacttgaac
        v  l  p   q  i  q   e  t  t  a   r  i  i   f  n  g   k  d  l  n

1561  cttgtcgaac gtagaattgc cgccgtgaac cccagtgatc cactcgagac gactaaaccg
        l  v  e   r  r  i   a  a  v  n   p  s  d   p  l  e   t  t  k  p

1621  gatatgacac tgaaagaggc tctgaagatt gccttcggat tcaacgaacc taatggcaat
        d  m  t   l  k  e   a  l  k  i   a  f  g   f  n  e   p  n  g  n

1681  ttgcagtatc aggggaaaga catcacagag tttgatttca atttcgatca gcagacttcc
        l  q  y   q  g  k   d  i  t  e   f  d  f   n  f  d   q  q  t  s

1741  caaaatatca aaaatcagtt ggcagagctg aatgccacca atatctacac ggttctcgat
        q  n  i   k  n  q   l  a  e  l   n  a  t   n  i  y   t  v  l  d

1801  aaaatcaaac ttaacgccaa gatgaacata ttgattcgag acaaacgctt ccactacgac
        k  i  k   l  n  a   k  m  n  i   l  i  r   d  k  r   f  h  y  d

1861  cgcaacaata tagccgtagg cgctgatgag tctgtcgtca aggaggctca tagggaagtt
        r  n  n   i  a  v   g  a  d  e   s  v  v   k  e  a   h  r  e  v

1921  atcaacagca gtactgaagg gctgttactt aatatcgaca aggacattcg gaagatcctg
        i  n  s   s  t  e   g  l  l  l   n  i  d   k  d  i   r  k  i  l

1981  tccgggtata tcgtggagat cgaggatacc gagggcctga aggaagtcat taacgaccgc
        s  g  y   i  v  e   i  e  d  t   e  g  l   k  e  v   i  n  d  r

2041  tatgatatgc tgaacatttc cagcttacga caggacggta agacatttat tgactttaaa
        y  d  m   l  n  i   s  s  l  r   q  d  g   k  t  f   i  d  f  k

2101  aagtataacg acaagctacc cctgtacatt tccaacccaa attacaaagt taatgtgtat
        k  y  n   d  k  l   p  l  y  i   s  n  p   n  y  k   v  n  v  y
```

Figure 3C

```
2161  gctgtaacca aggagaacac aatcatcaat ccaagcgaga acggcgatac cagcacaaat
         a  v  t   k  e  n   t  i  i   n  p  s   e  n  g   d  t  s   t  n

2221  ggaatcaaaa agatccttat atttagtaaa aaaggctacg agatcggttg aggatcc
         g  i  k   k  i  l   i  f  s   k  k  g   y  e  i   g  -
```

Figure 4A

```
   1  gatatcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga
                        m  d   a  m  k  r   g  l  c   c  v  l   l  l  c  g

  61  gcagtcttcg tttcgcccag cgccggcggg catggggacg ttggcatgca tgtgaaagaa
       a  v  f   v  s  p   s  a  g  g   h  g  d   v  g  m   h  v  k  e

 121  aaggagaaaa acaaggacga aaacaagcgt aaagacgaag aacgtaataa aacacaggag
       k  e  k   n  k  d   e  n  k  r   k  d  e   e  r  n   k  t  q  e

 181  gaacacttaa aggagatcat gaagcacata gtaaagattg aggtaaaagg cgaagaggct
       e  h  l   k  e  i   m  k  h  i   v  k  i   e  v  k   g  e  e  a

 241  gtaaagaagg aggcagcaga aaaactgttg gagaaggtgc cttctgacgt cttagagatg
       v  k  k   e  a  a   e  k  l  l   e  k  v   p  s  d   v  l  e  m

 301  tataaggcca tcggcggtaa gatctatatc gtggacggag acatcactaa acacatatct
       y  k  a   i  g  g   k  i  y  i   v  d  g   d  i  t   k  h  i  s

 361  ctcgaagctc tctccgagga caagaaaaag attaaagaca tctacgggaa ggatgcctta
       l  e  a   l  s  e   d  k  k  k   i  k  d   i  y  g   k  d  a  l

 421  ttgcacgagc actacgttta cgcaaaggag ggctatgagc ccgtgctcgt tattcagagt
       l  h  e   h  y  v   y  a  k  e   g  y  e   p  v  l   v  i  q  s

 481  agtgaggact acgtcgagaa taccgagaaa gctctgaatg tgtattacga gatcggaaag
       s  e  d   y  v  e   n  t  e  k   a  l  n   v  y  y   e  i  g  k

 541  attctgtccc gggacatcct gtccaaaatc aaccagccat accagaaatt ccttgatgtt
       i  l  s   r  d  i   l  s  k  i   n  q  p   y  q  k   f  l  d  v

 601  cttaacacaa tcaaaaacgc gtcagatagc gacgggcagg atcttctgtt tacaaatcaa
       l  n  t   i  k  n   a  s  d  s   d  g  q   d  l  l   f  t  n  q

 661  ctcaaggaac accccactga tttcagcgtg gagttcctcg agcagaattc taacgaagtc
       l  k  e   h  p  t   d  f  s  v   e  f  l   e  q  n   s  n  e  v

 721  caggaggtgt tcgccaaggc atttgcgtac tatatcgaac cccagcatcg cgatgtgctc
       q  e  v   f  a  k   a  f  a  y   y  i  e   p  q  h   r  d  v  l

 781  cagctgtacg ccccggaggc atttaactac atggacaaat tcaatgaaca ggagattaat
       q  l  y   a  p  e   a  f  n  y   m  d  k   f  n  e   q  e  i  n

 841  ctgtctctgg aggaactgaa agaccagagg atgctctccc ggtatgaaaa gtgggaaaag
       l  s  l   e  e  l   k  d  q  r   m  l  s   r  y  e   k  w  e  k

 901  atcaaacagc attaccagca ttggtccgac tccctgtcag aagaggggcg cggcctgttg
       i  k  q   h  y  q   h  w  s  d   s  l  s   e  e  g   r  g  l  l

 961  aaaaagttgc agattcccat cgagcctaag aaagatgata taatacactc tctaagccag
       k  k  l   q  i  p   i  e  p  k   k  d  d   i  i  h   s  l  s  q

1021  gaggagaagg aactcctgaa gcggatacaa atcgactcat ccgatttcct tagcacagaa
       e  e  k   e  l  l   k  r  i  q   i  d  s   s  d  f   l  s  t  e
```

Figure 4B

```
1081  gagaaggagt ttctaaaaaa acttcagata gatattagag attcactgag cgaggaagag
         e  k  e    f  l  k    k  l  q  i   d  i  r   d  s  l   s  e  e  e

1141  aaggagctgc tcaaccgaat tcaagtcgat agttcgaacc ccttgtcaga aaaagagaag
         k  e  l    l  n  r    i  q  v  d   s  s  n   p  l  s   e  k  e  k

1201  gaattcctga aaaagttgaa gctcgacatc cagccgtacg atattaatca gcggctacaa
         e  f  l    k  k  l    k  l  d  i   q  p  y   d  i  n   q  r  l  q

1261  gacaccggcg gtctgattga tagccccagc atcaaccttg acgtacggaa gcaatataag
         d  t  g    g  l  i    d  s  p  s   i  n  l   d  v  r   k  q  y  k

1321  cgcgacattc aaaatatcga cgccctatta catcaatcca taggctccac gctatacaat
         r  d  i    q  n  i    d  a  l  l   h  q  s   i  g  s   t  l  y  n

1381  aaaatctatc tatacgaaaa catgaatatt aacaatctca ccgctacact gggagcggac
         k  i  y    l  y  e    n  m  n  i   n  n  l   t  a  t   l  g  a  d

1441  ctggtcgata gtacagacaa cacaaagata aacagaggta ttttcaacga attcaaaaag
         l  v  d    s  t  d    n  t  k  i   n  r  g   i  f  n   e  f  k  k

1501  aactttaagt attcgatcag cagtaactat atgattgttg acatcaatga acggcccgca
         n  f  k    y  s  i    s  s  n  y   m  i  v   d  i  n   e  r  p  a

1561  ttagacaatg agaggttgaa gtggagaatt caactgagtc ctgatactag ggccggctat
         l  d  n    e  r  l    k  w  r  i   q  l  s   p  d  t   r  a  g  y

1621  ctggagaacg ggaaactgat cttacagcga aacatcgggc tggagatcaa ggatgtgcag
         l  e  n    g  k  l    i  l  q  r   n  i  g   l  e  i   k  d  v  q

1681  attatcaagc agagcgaaaa agaatacatt cgcatcgacg ccaaggtggt gcctaagtca
         i  i  k    q  s  e    k  e  y  i   r  i  d   a  k  v   v  p  k  s

1741  aagatcgata ccaagatcca ggaagctcag ctcaacatta accaggagtg gaataaagct
         k  i  d    t  k  i    q  e  a  q   l  n  i   n  q  e   w  n  k  a

1801  cttggtctgc caaaatacac caaacttatc acctttaatg tgcacaacag gtatgcctct
         l  g  l    p  k  y    t  k  l  i   t  f  n   v  h  n   r  y  a  s

1861  aatatcgtcg agtcagcata cctgattctc aatgaatgga agaacaatat tcagtctgac
         n  i  v    e  s  a    y  l  i  l   n  e  w   k  n  n   i  q  s  d

1921  ctgatcaaga aggtcacgaa ttatctggtg gacggaaatg gcagattcgt gttcaccgac
         l  i  k    k  v  t    n  y  l  v   d  g  n   g  r  f   v  f  t  d

1981  ataactttgc caaacattgc cgagcaatac actcatcagg atgaaattta cgagcaagtc
         i  t  l    p  n  i    a  e  q  y   t  h  q   d  e  i   y  e  q  v

2041  cactccaaag gtctgtatgt tccagagtca agatcgattc tgctccatgg tccatccaaa
         h  s  k    g  l  y    v  p  e  s   r  s  i   l  l  h   g  p  s  k

2101  ggggttgagc ttcgaaacga ttctgaggga tttatcgctg actttggagc cgctgtggat
         g  v  e    l  r  n    d  s  e  g   f  i  a   d  f  g   a  a  v  d
```

Figure 4C

```
2161  gactacgccg gatacctgtt ggataagaat cagtctgatc tcgtgacaaa tagcaaaaaa
        d  y  a   g  y  l   l  d  k   n  q  s   d  l  v   t  n  s  k  k

2221  ttcatagata ttttcaagga ggaagggagt aacctgactt cctatggccg cacgaacgag
        f  i  d   i  f  k   e  è  g   s  n  l   t  s  y   g  r  t  n  e

2281  gctgaatttt ttgcggaagc ctttagactt atgcacagca ccgaccatgc tgaaaggttg
        a  e  f   f  a  e   a  f  r   l  m  h   s  t  d   h  a  e  r  l

2341  aaggtgcaaa agaatgcccc taaaaccttc cagttcataa atgaccagat caagttcatc
        k  v  q   k  n  a   p  k  t   f  q  f   i  n  d   q  i  k  f  i

2401  atcaactctt gaggatcc
        i  n  s   -
```

Figure 5A

```
   1 gatatcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga
                          m  d   a  m  k  r  g  l  c  c  v  l  l  c  g

  61 gcagtcttcg tttcgcccag cgccggcggg catggggacg ttggcatgca tgtgaaagaa
      a  v  f  v  s  p  s  a  g  g  h  g  d  v  g  m  h  v  k  e

 121 aaggagaaaa acaaggacga aaacaagcgt aaagacgaag aacgtaataa aacacaggag
      k  e  k  n  k  d  e  n  k  r  k  d  e  e  r  n  k  t  q  e

 181 gaacacttaa aggagatcat gaagcacata gtaaagattg aggtaaaagg cgaagaggct
      e  h  l  k  e  i  m  k  h  i  v  k  i  e  v  k  g  e  e  a

 241 gtaaagaagg aggcagcaga aaaactgttg gagaaggtgc cttctgacgt cttagagatg
      v  k  k  e  a  a  e  k  l  l  e  k  v  p  s  d  v  l  e  m

 301 tataaggcca tcggcggtaa gatctatatc gtggacggag acatcactaa acacatatct
      y  k  a  i  g  g  k  i  y  i  v  d  g  d  i  t  k  h  i  s

 361 ctcgaagctc tctccgagga caagaaaaag attaaagaca tctacgggaa ggatgcctta
      l  e  a  l  s  e  d  k  k  k  i  k  d  i  y  g  k  d  a  l

 421 ttgcacgagc actacgttta cgcaaaggag ggctatgagc ccgtgctcgt tattcagagt
      l  h  e  h  y  v  y  a  k  e  g  y  e  p  v  l  v  i  q  s

 481 agtgaggact acgtcgagaa taccgagaaa gctctgaatg tgtattacga gatcggaaag
      s  e  d  y  v  e  n  t  e  k  a  l  n  v  y  y  e  i  g  k

 541 attctgtccc gggacatcct gtccaaaatc aaccagccat accagaaatt ccttgatgtt
      i  l  s  r  d  i  l  s  k  i  n  q  p  y  q  k  f  l  d  v

 601 cttaacacaa tcaaaaacgc gtcagatagc gacgggcagg atcttctgtt tacaaatcaa
      l  n  t  i  k  n  a  s  d  s  d  g  q  d  l  l  f  t  n  q

 661 ctcaaggaac accccactga tttcagcgtg gagttcctcg agcagaattc taacgaagtc
      l  k  e  h  p  t  d  f  s  v  e  f  l  e  q  n  s  n  e  v

 721 caggaggtgt tcgccaaggc atttgcgtac tatatcgaac cccagcatcg cgatgtgctc
      q  e  v  f  a  k  a  f  a  y  y  i  e  p  q  h  r  d  v  l

 781 cagctgtacg ccccggaggc atttaactac atggacaaat tcaatgaaca ggagattaat
      q  l  y  a  p  e  a  f  n  y  m  d  k  f  n  e  q  e  i  n

 841 ctgtctctgg aggaactgaa agaccagagg atgctctccc ggtatgaaaa gtgggaaaag
      l  s  l  e  e  l  k  d  q  r  m  l  s  r  y  e  k  w  e  k

 901 atcaaacagc attaccagca ttggtccgac tccctgtcag aagaggggcg cggcctgttg
      i  k  q  h  y  q  h  w  s  d  s  l  s  e  e  g  r  g  l  l

 961 aaaaagttgc agattcccat cgagcctaag aaagatgata taatacactc tctaagccag
      k  k  l  q  i  p  i  e  p  k  k  d  d  i  i  h  s  l  s  q

1021 gaggagaagg aactcctgaa gcggatacaa atcgactcat ccgatttcct tagcacagaa
      e  e  k  e  l  l  k  r  i  q  i  d  s  s  d  f  l  s  t  e
```

Figure 5B

```
1081  gagaaggagt ttctaaaaaa acttcagata gatattagag attcactgag cgaggaagag
         e  k  e   f  l  k   k  l  q  i   d  i  r   d  s  l   s  e  e  e

1141  aaggagctgc tcaaccgaat tcaagtcgat agttcgaacc ccttgtcaga aaaagagaag
         k  e  l   l  n  r   i  q  v  d   s  s  n   p  l  s   e  k  e  k

1201  gaattcctga aaaagttgaa gctcgacatc cagccgtacg atattaatca gcggctacaa
         e  f  l   k  k  l   k  l  d  i   q  p  y   d  i  n   q  r  l  q

1261  gacaccggcg gtctgattga tagccccagc atcaaccttg acgtacggaa gcaatataag
         d  t  g   g  l  i   d  s  p  s   i  n  l   d  v  r   k  q  y  k

1321  cgcgacattc aaaatatcga cgccctatta catcaatcca taggctccac gctatacaat
         r  d  i   q  n  i   d  a  l  l   h  q  s   i  g  s   t  l  y  n

1381  aaaatctatc tatacgaaaa catgaatatt aacaatctca ccgctacact gggagcggac
         k  i  y   l  y  e   n  m  n  i   n  n  l   t  a  t   l  g  a  d

1441  ctggtcgata gtacagacaa cacaaagata aacagaggta ttttcaacga attcaaaaag
         l  v  d   s  t  d   n  t  k  i   n  r  g   i  f  n   e  f  k  k

1501  aactttaagt attcgatcag cagtaactat atgattgttg acatcaatga acggcccgca
         n  f  k   y  s  i   s  s  n  y   m  i  v   d  i  n   e  r  p  a

1561  ttagacaatg agaggttgaa gtggagaatt caactgagtc ctgatactag ggccggctat
         l  d  n   e  r  l   k  w  r  i   q  l  s   p  d  t   r  a  g  y

1621  ctggagaacg ggaaactgat cttacagcga aacatcgggc tggagatcaa ggatgtgcag
         l  e  n   g  k  l   i  l  q  r   n  i  g   l  e  i   k  d  v  q

1681  attatcaagc agagcgaaaa agaatacatt cgcatcgacg ccaaggtggt gtagggatcc
         i  i  k   q  s  e   k  e  y  i   r  i  d   a  k  v   v  -
```

Figure 6

```
  1  gatatcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga
                    m  d   a  m  k  r   g  l  c   c  v  l   l  l  c  g

 61  gcagtcttcg tttcgcccag cgccggcggg catggggacg ttggcatgca tgtgaaagaa
       a  v  f   v  s  p   s  a  g  g   h  g  d   v  g  m   h  v  k  e

121  aaggagaaaa acaaggacga aaacaagcgt aaagacgaag aacgtaataa aacacaggag
       k  e  k   n  k  d   e  n  k  r   k  d  e   e  r  n   k  t  q  e

181  gaacacttaa aggagatcat gaagcacata gtaaagattg aggtaaaagg cgaagaggct
       e  h  l   k  e  i   m  k  h  i   v  k  i   e  v  k   g  e  e  a

241  gtaaagaagg aggcagcaga aaaactgttg gagaaggtgc cttctgacgt cttagagatg
       v  k  k   e  a  a   e  k  l  l   e  k  v   p  s  d   v  l  e  m

301  tataaggcca tcggcggtaa gatctatatc gtggacggag acatcactaa acacatatct
       y  k  a   i  g  g   k  i  y  i   v  d  g   d  i  t   k  h  i  s

361  ctcgaagctc tctccgagga caagaaaaag attaaagaca tctacgggaa ggatgcctta
       l  e  a   l  s  e   d  k  k  k   i  k  d   i  y  g   k  d  a  l

421  ttgcacgagc actacgttta cgcaaaggag ggctatgagc ccgtgctcgt tattcagagt
       l  h  e   h  y  v   y  a  k  e   g  y  e   p  v  l   v  i  q  s

481  agtgaggact acgtcgagaa taccgagaaa gctctgaatg tgtattacga gatcggaaag
       s  e  d   y  v  e   n  t  e  k   a  l  n   v  y  y   e  i  g  k

541  attctgtccc gggacatcct gtccaaaatc aaccagccat accagaaatt ccttgatgtt
       i  l  s   r  d  i   l  s  k  i   n  q  p   y  q  k   f  l  d  v

601  cttaacacaa tcaaaaacgc gtcagatagc gacgggcagg atcttctgtt tacaaatcaa
       l  n  t   i  k  n   a  s  d  s   d  g  q   d  l  l   f  t  n  q

661  ctcaaggaac accccactga tttcagcgtg gagttcctcg agcagaattc taacgaagtc
       l  k  e   h  p  t   d  f  s  v   e  f  l   e  q  n   s  n  e  v

721  caggaggtgt tcgccaaggc attttgagga tcc
       q  e  v   f  a  k   a  f  -
```

Figure 7A

```
   1  gatatcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga
                         m  d   a  m  k  r   g  l  c   c  v  l   l  l  c  g

  61  gcagtcttcg tttcgcccag cagcgctggg ccaactgtgc ccgacagaga caatgatgga
       a  v  f   v  s  p   s  s  a  g   p  t  v   p  d  r   d  n  d  g

 121  atccctgata gtctagaggt tgagggatac acggtagatg tcaagaacaa aaggactttt
       i  p  d   s  l  e   v  e  g  y   t  v  d   v  k  n   k  r  t  f

 181  ctctcgcctt ggatctcaaa tatccatgag aagaaggggc ttaccaagta caagtcctcc
       l  s  p   w  i  s   n  i  h  e   k  k  g   l  t  k   y  k  s  s

 241  cccgagaagt ggtctaccgc ttccgatcca tatagcgatt tcgagaaggt cacaggccgg
       p  e  k   w  s  t   a  s  d  p   y  s  d   f  e  k   v  t  g  r

 301  atcgataaac aggtgtctcc agaggctaga cacccccctgg tagcagccta cccgattgta
       i  d  k   q  v  s   p  e  a  r   h  p  l   v  a  a   y  p  i  v

 361  cacgtggaca tggagaacat cattctaagc aaaaacgagg accagtccac acaaaacact
       h  v  d   m  e  n   i  i  l  s   k  n  e   d  q  s   t  q  n  t

 421  gactccgaga cccgcaccat atctaaacag accagtactt caaggaccca cacctctgaa
       d  s  e   t  r  t   i  s  k  q   t  s  t   s  r  t   h  t  s  e

 481  gtgcacggca atgcggaagt ccatgcatcg tttttcgata ttggtggctc cgtgtcagcc
       v  h  g   n  a  e   v  h  a  s   f  f  d   i  g  g   s  v  s  a

 541  ggctttagca atagccagtc ctcgacggtt gccattgacc actcactgtc attagcaggt
       g  f  s   n  s  q   s  s  t  v   a  i  d   h  s  l   s  l  a  g

 601  gagaggactt gggctgaaac tatgggtctg aataccgccg atacggcccg gctcaacgca
       e  r  t   w  a  e   t  m  g  l   n  t  a   d  t  a   r  l  n  a

 661  aatattcggt acgtcaacac agggactgct cctatatata acgtgctgcc tacgacaagt
       n  i  r   y  v  n   t  g  t  a   p  i  y   n  v  l   p  t  t  s

 721  cttgtcctgg gcaaacagca gaccctcgca accattaagg caaaggaaaa tcagctgagc
       l  v  l   g  k  q   q  t  l  a   t  i  k   a  k  e   n  q  l  s

 781  cagatcctcg cccctaacaa ctattatcca tccaaaaatt tagcccccat agccctgaac
       q  i  l   a  p  n   n  y  y  p   s  k  n   l  a  p   i  a  l  n

 841  gcccaggacg acttttcctc tacccccata actatgaatt acaatcagtt cctggagctg
       a  q  d   d  f  s   s  t  p  i   t  m  n   y  n  q   f  l  e  l

 901  gaaaagacga agcagctgag actagacacc gatcaggtgt atggaaacat agcgacatat
       e  k  t   k  q  l   r  l  d  t   d  q  v   y  g  n   i  a  t  y

 961  aactttgaga acggccgcgt gcgcgtcgac actgggtcac agtggtctga agttctgccg
       n  f  e   n  g  r   v  r  v  d   t  g  s   q  w  s   e  v  l  p

1021  caaattcaag agacaaccgc cagaattatc tttaatggga aggacttgaa ccttgtcgaa
       q  i  q   e  t  t   a  r  i  i   f  n  g   k  d  l   n  l  v  e
```

Figure 7B

```
1081  cgtagaattg ccgccgtgca gcccagtgat ccactcgaga cgactaaacc ggatatgaca
        r  r  i   a  a  v   q  p  s  d   p  l  e   t  t  k   p  d  m  t

1141  ctgaaagagg ctctgaagat tgccttcgga ttcaacgaac ctaatggcaa tttgcagtat
        l  k  e   a  l  k   i  a  f  g   f  n  e   p  n  g   n  l  q  y

1201  caggggaaag acatcacaga gtttgatttc aatttcgatc agcagacttc ccaaaatatc
        q  g  k   d  i  t   e  f  d  f   n  f  d   q  q  t   s  q  n  i

1261  aaaaatcagt tggcagagct gcaggccacc aatatctaca cggttctcga taaaatcaaa
        k  n  q   l  a  e   l  q  a  t   n  i  y   t  v  l   d  k  i  k

1321  cttaacgcca agatgaacat attgattcga gacaaacgct tccactacga ccgcaacaat
        l  n  a   k  m  n   i  l  i  r   d  k  r   f  h  y   d  r  n  n

1381  atagccgtag gcgctgatga gtctgtcgtc aaggaggctc atagggaagt tatccagagc
        i  a  v   g  a  d   e  s  v  v   k  e  a   h  r  e   v  i  q  s

1441  agtactgaag ggctgttact taatatcgac aaggacattc ggaagatcct gtccgggtat
        s  t  e   g  l  l   l  n  i  d   k  d  i   r  k  i   l  s  g  y

1501  atcgtggaga tcgaggatac cgagggcctg aaggaagtca ttaacgaccg ctatgatatg
        i  v  e   i  e  d   t  e  g  l   k  e  v   i  n  d   r  y  d  m

1561  ctgcagattt ccagcttacg acaggacggt aagacattta ttgactttaa aaagtataac
        l  q  i   s  s  l   r  q  d  g   k  t  f   i  d  f   k  k  y  n

1621  gacaagctac ccctgtacat ttccaaccca aattacaaag ttaatgtgta tgctgtaacc
        d  k  l   p  l  y   i  s  n  p   n  y  k   v  n  v   y  a  v  t

1681  aaggagaaca caatcatcca gccaagcgag aacggcgata ccagcacaaa tggaatcaaa
        k  e  n   t  i  i   q  p  s  e   n  g  d   t  s  t   n  g  i  k

1741  aagatcctta tatttagtaa aaaaggctac gagatcggtt gaggatcc
        k  i  l   i  f  s   k  k  g  y   e  i  g   -
```

Figure 8A

```
   1  gatatcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga
                   m  d   a  m  k  r   g  l  c   c  v  l   l  l  c  g

  61  gcagtcttcg tttcgcccag cgccggcggg catggggacg ttggcatgca tgtgaaagaa
       a  v  f   v  s  p   s  a  g  g   h  g  d   v  g  m   h  v  k  e

 121  aaggagaaaa acaaggacga aaacaagcgt aaagacgaag aacgtcagaa aacacaggag
       k  e  k   n  k  d   e  n  k  r   k  d  e   e  r  q   k  t  q  e

 181  gaacacttaa aggagatcat gaagcacata gtaaagattg aggtaaaagg cgaagaggct
       e  h  l   k  e  i   m  k  h  i   v  k  i   e  v  k   g  e  e  a

 241  gtaaagaagg aggcagcaga aaaactgttg gagaaggtgc cttctgacgt cttagagatg
       v  k  k   e  a  a   e  k  l  l   e  k  v   p  s  d   v  l  e  m

 301  tataaggcca tcggcggtaa gatctatatc gtggacggag acatcactaa acacatatct
       y  k  a   i  g  g   k  i  y  i   v  d  g   d  i  t   k  h  i  s

 361  ctcgaagctc tctccgagga caagaaaaag attaaagaca tctacgggaa ggatgcctta
       l  e  a   l  s  e   d  k  k  k   i  k  d   i  y  g   k  d  a  l

 421  ttgcacgagc actacgttta cgcaaaggag ggctatgagc ccgtgctcgt tattcagagt
       l  h  e   h  y  v   y  a  k  e   g  y  e   p  v  l   v  i  q  s

 481  agtgaggact acgtcgagaa taccgagaaa gctctgaatg tgtattacga gatcggaaag
       s  e  d   y  v  e   n  t  e  k   a  l  n   v  y  y   e  i  g  k

 541  attctgtccc gggacatcct gtccaaaatc aaccagccat accagaaatt ccttgatgtt
       i  l  s   r  d  i   l  s  k  i   n  q  p   y  q  k   f  l  d  v

 601  cttaacacaa tcaaacaggc gtcagatagc gacgggcagg atcttctgtt tacaaatcaa
       l  n  t   i  k  q   a  s  d  s   d  g  q   d  l  l   f  t  n  q

 661  ctcaaggaac accccactga tttcagcgtg gagttcctcg agcagaattc taacgaagtc
       l  k  e   h  p  t   d  f  s  v   e  f  l   e  q  n   s  n  e  v

 721  caggaggtgt tcgccaaggc atttgcgtac tatatcgaac cccagcatcg cgatgtgctc
       q  e  v   f  a  k   a  f  a  y   y  i  e   p  q  h   r  d  v  l

 781  cagctgtacg ccccggaggc atttaactac atggacaaat tcaatgaaca ggagattcag
       q  l  y   a  p  e   a  f  n  y   m  d  k   f  n  e   q  e  i  q

 841  ctgtctctgg aggaactgaa agaccagagg atgctctccc ggtatgaaaa gtgggaaaag
       l  s  l   e  e  l   k  d  q  r   m  l  s   r  y  e   k  w  e  k

 901  atcaaacagc attaccagca ttggtccgac tccctgtcag aagaggggcg cggcctgttg
       i  k  q   h  y  q   h  w  s  d   s  l  s   e  e  g   r  g  l  l

 961  aaaaagttgc agattcccat cgagcctaag aaagatgata taatacactc tctaagccag
       k  k  l   q  i  p   i  e  p  k   k  d  d   i  i  h   s  l  s  q

1021  gaggagaagg aactcctgaa gcggatacaa atcgactcat ccgatttcct tagcacagaa
       e  e  k   e  l  l   k  r  i  q   i  d  s   s  d  f   l  s  t  e
```

Figure 8B

```
1081  gagaaggagt ttctaaaaaa acttcagata gatattagag attcactgag cgaggaagag
        e  k  e   f  l  k   k  l  q  i   d  i  r   d  s  l   s  e  e  e

1141  aaggagctgc tcaaccgaat tcaagtcgat agttcgaacc ccttgtcaga aaaagagaag
        k  e  l   l  n  r   i  q  v  d   s  s  n   p  l  s   e  k  e  k

1201  gaattcctga aaaagttgaa gctcgacatc cagccgtacg atattaatca gcggctacaa
        e  f  l   k  k  l   k  l  d  i   q  p  y   d  i  n   q  r  l  q

1261  gacaccggcg gtctgattga tagccccagc atcaaccttg acgtacggaa gcaatataag
        d  t  g   g  l  i   d  s  p  s   i  n  l   d  v  r   k  q  y  k

1321  cgcgacattc aaaatatcga cgccctatta catcaatcca taggctccac gctatacaat
        r  d  i   q  n  i   d  a  l  l   h  q  s   i  g  s   t  l  y  n

1381  aaaatctatc tatacgaaaa catgaatatt aaccagctca ccgctacact gggagcggac
        k  i  y   l  y  e   n  m  n  i   n  q  l   t  a  t   l  g  a  d

1441  ctggtcgata gtacagacaa cacaaagata aacagaggta ttttcaacga attcaaaaag
        l  v  d   s  t  d   n  t  k  i   n  r  g   i  f  n   e  f  k  k

1501  aactttaagt attcgatcag cagtaactat atgattgttg acatcaatga acggcccgca
        n  f  k   y  s  i   s  s  n  y   m  i  v   d  i  n   e  r  p  a

1561  ttagacaatg agaggttgaa gtggagaatt caactgagtc ctgatactag ggccggctat
        l  d  n   e  r  l   k  w  r  i   q  l  s   p  d  t   r  a  g  y

1621  ctggagaacg ggaaactgat cttacagcga aacatcgggc tggagatcaa ggatgtgcag
        l  e  n   g  k  l   i  l  q  r   n  i  g   l  e  i   k  d  v  q

1681  attatcaagc agagcgaaaa agaatacatt cgcatcgacg ccaaggtggt gcctaagtca
        i  i  k   q  s  e   k  e  y  i   r  i  d   a  k  v   v  p  k  s

1741  aagatcgata ccaagatcca ggaagctcag ctcaacatta accaggagtg gaataaagct
        k  i  d   t  k  i   q  e  a  q   l  n  i   n  q  e   w  n  k  a

1801  cttggtctgc caaaatacac caaacttatc acctttaatg tgcacaacag gtatgcctct
        l  g  l   p  k  y   t  k  l  i   t  f  n   v  h  n   r  y  a  s

1861  aatatcgtcg agtcagcata cctgattctc aatgaatgga agaacaatat tcagtctgac
        n  i  v   e  s  a   y  l  i  l   n  e  w   k  n  n   i  q  s  d

1921  ctgatcaaga aggtcacgaa ttatctggtg gacggaaatg gcagattcgt gttcaccgac
        l  i  k   k  v  t   n  y  l  v   d  g  n   g  r  f   v  f  t  d

1981  ataactttgc caaacattgc cgagcaatac actcatcagg atgaaattta cgagcaagtc
        i  t  l   p  n  i   a  e  q  y   t  h  q   d  e  i   y  e  q  v

2041  cactccaaag gtctgtatgt tccagagtca agatcgattc tgctccatgg tccatccaaa
        h  s  k   g  l  y   v  p  e  s   r  s  i   l  l  h   g  p  s  k

2101  ggggttgagc ttcgacagga ttctgaggga tttatcgctg actttggagc cgctgtggat
        g  v  e   l  r  q   d  s  e  g   f  i  a   d  f  g   a  a  v  d
```

Figure 8C

```
2161  gactacgccg gatacctgtt ggataagcag cagtctgatc tcgtgacaaa tagcaaaaaa
        d  y  a   g  y  l   l  d  k  q   q  s  d   l  v  t   n  s  k  k

2221  ttcatagata ttttcaagga ggaagggagt cagctgactt cctatggccg cacgaacgag
        f  i  d   i  f  k   e  e  g  s   q  l  t   s  y  g   r  t  n  e

2281  gctgaatttt ttgcggaagc ctttagactt atgcacagca ccgaccatgc tgaaaggttg
        a  e  f   f  a  e   a  f  r  l   m  h  s   t  d  h   a  e  r  l

2341  aaggtgcaaa agaatgcccc taaaaccttc cagttcataa atgaccagat caagttcatc
        k  v  q   k  n  a   p  k  t  f   q  f  i   n  d  q   i  k  f  i

2401  atcaactctt gaggatcc
        i  n  s   -
```

Figure 9A

```
TPA-Human PA    82 agcgctgggccaactgtgcccgacagagacaatgatggaatccctgatagtctagaggttgagggataca
B.a.- PA       595 ..t.....a..t..g..t..a...c.t...................tcat.......a..a.....t.

TPA-Human PA   152 cggtagatgtcaagaacaaaaggacttttctctcgccttggatctcaaatatccatgagaagaaggggct
B.a.- PA       665 ....t.......a..t....a........t..a..a.....t..t.....t.....a.....a..at.

TPA-Human PA   222 taccaagtacaagtcctcccccgagaagtggtctaccgcttccgatccatatagcgatttcgagaaggtc
B.a.- PA       735 a.....a..t..a..a..t..t..a..a...agc..g.....t.....g..c..t........a.....t

TPA-Human PA   292 acaggccggatcgataaaaatgtgtctccagaggctagacacccccctggtagcagcctacccgattgtac
B.a.- PA       805 .....a.....t.....g.....a..a........a...........t..g.....t..t..........

TPA-Human PA   362 acgtggacatggagaacatcattctaagcaaaaacgaggaccagtccacacaaaacactgactccgagac
B.a.- PA       875 .t..a..t........t..t.....ctca.....t.....t..a.........g..t.....tagt..a..

TPA-Human PA   432 ccgcaccatatctaaaaacaccagtacttcaaggacccacacctctgaagtgcacggcaatgcggaagtc
B.a.- PA       945 ga.a..a...ag......t..ttc...aagt.....a..t..tag......a..t..a.....a.....g

TPA-Human PA   502 catgcatcgttttcgatattggtggctccgtgtcagccggctttagcaatagcaactcctcgacggttg
B.a.- PA      1015 .....g.....c..t...........gagt..a..t..a..a......t...tcg..t..aagt.....c.

TPA-Human PA   572 ccattgaccactcactgtcattagcaggtgagaggacttgggctgaaactatgggtctgaataccgccga
B.a.- PA      1085 .a.....t..t.....a..tc.......g..a..a...............a......t.a........t..

TPA-Human PA   642 tacggcccggctcaacgcaaatattcggtacgtcaacacagggactgctcctatatataacgtgctgcct
B.a.- PA      1155 ...a..aa.at.a..t..c......a.a..t..a..t..t.....g.....a..c..c......t.a..a

TPA-Human PA   712 acgacaagtcttgtcctgggcaaaaatcagaccctcgcaaccattaaggcaaaggaaaatcagctgagcc
B.a.- PA      1225 .....ttcgt.a..gt.a..a........a..a.....g..a......a..t........c..at.a..t.

TPA-Human PA   782 agatcctcgcccctaacaactattatccatccaaaaatttagcccccatagccctgaacgcccaggacga
B.a.- PA      1295 .a..a..t..a.....t..t........t..t.....c..g..g..a..c..at.a..t..a..a.....

TPA-Human PA   852 cttttcctctacccccataactatgaattacaatcagttcctggagctggaaaagacgaagcagctgaga
B.a.- PA      1365 t..cagt.....t..a..t..a..............a..t..t...t.a.....a.....a..at.a...

TPA-Human PA   922 ctagacaccgatcaggtgtatggaaacatagcgacatataactttgagaacggccgcgtgcgcgtcgaca
B.a.- PA      1435 t....t..g.....a..a.....g..t.....a.....c..t.....a..t..aa.a...a.g..g..t.

TPA-Human PA   992 ctgggtcaaactggtctgaagttctgccgcaaattcaagagacaaccgccagaattatctttaatgggaa
B.a.- PA      1505 .a..c..g......ag......gt.a..............a......t..ac.t..c..t........a..

TPA-Human PA  1062 ggacttgaaccttgtcgaacgtagaattgccgccgtgaaccccagtgatccactcgagacgactaaaccg
B.a.- PA      1575 a..t..a..t..g..a....a.gc.g..a..g..g..t..t..t.........t.a..a............

TPA-Human PA  1132 gatatgacactgaaagaggctctgaagattgccttcggattcaacgaacctaatggcaatttgcagtatc
B.a.- PA      1645 .........t.a.....a..c..t..a..a..a..t.....t........g.....a..c..a..a....

TPA-Human PA  1202 aggggaaagacatcacagagtttgatttcaatttcgatcagcagacttcccaaaatatcaaaaatcagtt
B.a.- PA      1715 .a...........a..c..a........t...........a..a..a..t..........g........

TPA-Human PA  1272 ggcagagctgaatgccaccaatatctacacggttctcgataaaatcaaacttaacgccaagatgaacata
B.a.- PA      1785 a..g..at.a..c..a..t..c..a..t..t..at.a............t.a..t..a..a.....t..t

TPA-Human PA  1342 ttgattcgagacaaacgcttccactacgaccgcaacaatatagccgtaggcgctgatgagtctgtcgtca
B.a.- PA      1855 ..a..aa....t.....t..t..t..t..ta.a..t..c.....a..t..g..g........a..a..t.

TPA-Human PA  1412 aggaggctcatagggaagttatcaacagcagtactgaagggctgttacttaatatcgacaaggacattcg
B.a.- PA      1925 .............a.....a..t..ttcgtca..a..g..at.a..gt.a.....t..t.....t..aa.

TPA-Human PA  1482 gaagatcctgtccgggtatatcgtggagatcgaggataccgagggcctgaaggaagtcattaacgaccgc
B.a.- PA      1995 a..a..at.a..a..t.....t..a..a..t..a.....t..a..g..t..a.....t..a..t...a.a

TPA-Human PA  1552 tatgatatgctgaacatttccagcttacgacaggacggtaagacatttattgactttaaaaagtataacg
B.a.- PA      2065 .........t....t.....t..t.....g..a..t..a..a.........a..t........a.....t.
```

Figure 9B

```
TPA-Human PA    1622 acaagctacccctgtacatttccaacccaaattacaaagttaatgtgtatgctgtaaccaaggagaacac
B.a.- PA        2135 .t..at...gt.a..t..aagt..t..c.....t..g..a.....a........t..t..a..a.....

TPA-Human PA    1692 aatcatcaatccaagcgagaacggcgataccagcacaaatggaatcaaaaagatccttatatttagtaaa
B.a.- PA        2205 t..t..t.....t..t.....t..g.....t..t..c..c..g.....g..a..tt.a..c...tc....

TPA-Human PA    1762 aaaggctacgagatcggttga
B.a.- PA        2275 .......t.....a..a.a.
```

Figure 10A

```
TPA-human fu     82 gaagtgaagcaagaaaatcgacttctgaacgagagcgaaagttcatcacagggtcttctc
B.a.- PA         88 .....t..a..g..g..c..gt.at.a..t..atca...tcaagt..c.....gt.a..a

TPA-human fu    142 ggatactacttcagtgacttgaatttccaagcaccaatggtggtgactagtagcaccacc
B.a.- PA        148 .......t..t.....t.......t.......c........t..ctc.tct..t..a

TPA-human fu    202 ggcgatttgagcattcccagctctgagttggagaacattcccagcgaaaatcagtacttc
B.a.- PA        208 ..g.....atct.....t..t........a..a..t.....atcg.....c..a..t..t

TPA-human fu    262 cagtctgctatctggtccggattcattaaggttaaaaagtccgacgaatatacatttgct
B.a.- PA        268 ..a........t.....a.....t..c..a.....g...agt..t...............

TPA-human fu    322 acctcggcggataaccatgtgacaatgtgggtggacgaccaggaagtgatcaacaaggct
B.a.- PA        328 ..t..c..t.....t.....a...........a..t.....a........t..t..a...

TPA-human fu    382 tcaaactctaataaaatccggctcgagaaggggaggctctaccagatcaaaattcagtac
B.a.- PA        388 ..t..t.....c......a.at.a..a..a..a..at.a..t..a..a........a..t

TPA-human fu    442 cagcgggaaaaccctacagaaaaaggactcgatttcaagctgtactggacagatagccaa
B.a.- PA        448 ..a..a.....t.....t.........t.g.........t..........c...tct...

TPA-human fu    502 aacaagaaagaagttatcagctcagacaatctgcagttacccgagctcaagcagaagagt
B.a.- PA        508 ..t..a........g..ttctagt..t..ct.a..a..g..a..at.a..a..a..atc.

TPA-human fu    562 tctaa------------------tacaagcgctgggccaactgtgcccgacagagacaat
B.a.- PA        568 ..g..ctcaagaaaaaagcgaag......t.....a..t..g..t..a...c.t......

TPA-human fu    604 gatggaatccctgatagtctagaggttgagggatacacggtagatgtcaagaacaaaagg
B.a.- PA        628 ...............tcat.......a..a.....t.....t........a..t.....a

TPA-human fu    664 acttttctctcgccttggatctcaaatatccatgagaagaaggggcttaccaagtacaag
B.a.- PA        688 ........t..a..a.....t..t.....t.....a.....a..at.a.....a..t..a

TPA-human fu    724 tcctcccccgagaagtggtctaccgcttccgatccatatagcgatttcgagaaggtcaca
B.a.- PA        748 ..a..t..t..a..a...agc..g.....t.....g..c..t........a.....t...

TPA-human fu    784 ggccggatcgataaaaatgtgtctccagaggctagacacccccctggtagcagcctacccg
B.a.- PA        808 ..a.....t.....g.....a..a........a...........t..g.....t..t...

TPA-human fu    844 attgtacacgtggacatggagaacatcattctaagcaaaaacgaggaccagtccacacaa
B.a.- PA        868 ........t..a..t........t..t.....ctca.....t.....t..a........g

TPA-human fu    904 aacactgactccgagacccgcaccatatctaaaaacaccagtacttcaaggacccacacc
B.a.- PA        928 ..t.....tagt..a..ga.a..a...ag......t..ttc...aagt.....a..t..t

TPA-human fu    964 tctgaagtgcacggcaatgcggaagtccatgcatcgtttttcgatattggtggctccgtg
B.a.- PA        988 ag......a..t..a.....a.....g.....g.....c..t...........gagt..a

TPA-human fu   1024 tcagccggctttagcaatagcaactcctcgacggttgccattgaccactcactgtcatta
B.a.- PA       1048 ..t..a..a.....t...tcg..t..aagt.....c..a.....t..t.....a..tc..

TPA-human fu   1084 gcaggtgagaggacttgggctgaaactatgggtctgaataccgccgatacggcccggctc
B.a.- PA       1108 .....g..a..a................a......t.a.........t.....a..aa.at.a

TPA-human fu   1144 aacgcaaatattcggtacgtcaacacagggactgctcctatatataacgtgctgcctacg
B.a.- PA       1168 ..t..c......a.a..t..a..t..t.....g.....a..c..c......t.a..a...

TPA-human fu   1204 acaagtcttgtcctgggcaaaaatcagaccctcgcaaccattaaggcaaaggaaaatcag
B.a.- PA       1228 ..ttcgt.a..gt.a..a.........a..a.....g..a.....a..t........c..a

TPA-human fu   1264 ctgagccagatcctcgcccctaacaactattatccatccaaaaatttagcccccatagcc
B.a.- PA       1288 t.a..t..a..a..t..a.....t..t.........t..t.....c..g..g..a..c..a

TPA-human fu   1324 ctgaacgcccaggacgacttttcctctacccccataactatgaattacaatcagttcctg
B.a.- PA       1348 t.a..t..a..a.....t..cagt.....t..a..t..a..............a..t..t
```

Figure 10B

```
TPA-human fu    1384 gagctggaaaagacgaagcagctgagactagacaccgatcaggtgtatggaaacatagcg
B.a.- PA        1408 ...t.a.....a.....a..at.a...t....t..g.....a..a.....g..t.....a

TPA-human fu    1444 acatataactttgagaacggccgcgtgcgcgtcgacactgggtcaaactggtctgaagtt
B.a.- PA        1468 .....c..t.....a..t..aa.a...a.g..g..t..a..c..g......ag......g

TPA-human fu    1504 ctgccgcaaattcaagagacaaccgccagaattatctttaatgggaaggacttgaacctt
B.a.- PA        1528 t.a.............a.....t..ac.t..c..t........a..a..t..a..t..g

TPA-human fu    1564 gtcgaacgtagaattgccgccgtgaaccccagtgatccactcgagacgactaaaccggat
B.a.- PA        1588 ..a...a.gc.g..a..g..g..t..t..t.........t.a..a...............

TPA-human fu    1624 atgacactgaaagaggctctgaagattgccttcggattcaacgaacctaatggcaatttg
B.a.- PA        1648 ......t.a.....a..c..t..a..a..a..t.....t........g.....a..c..a

TPA-human fu    1684 cagtatcaggggaaagacatcacagagtttgatttcaatttcgatcagcagacttcccaa
B.a.- PA        1708 ..a.....a...........a..c..a.......t..........a..a..a..t...

TPA-human fu    1744 aatatcaaaaatcagttggcagagctgaatgccaccaatatctacacggttctcgataaa
B.a.- PA        1768 ........g........a..g..at.a..c..a..t..c..a..t..t..at.a......

TPA-human fu    1804 atcaaacttaacgccaagatgaacatattgattcgagacaaacgcttccactacgaccgc
B.a.- PA        1828 ......t.a..t..a..a.....t..t..a..aa....t.....t..t..t..t..ta.a

TPA-human fu    1864 aacaatatagccgtaggcgctgatgagtctgtcgtcaaggaggctcatagggaagttatc
B.a.- PA        1888 ..t..c.....a..t..g..g........a..a..t..............a.....a..t

TPA-human fu    1924 aacagcagtactgaagggctgttacttaatatcgacaaggacattcggaagatcctgtcc
B.a.- PA        1948 ..ttcgtca..a..g..at.a..gt.a.....t..t.....t..aa.a..a..at.a..a

TPA-human fu    1984 gggtatatcgtggagatcgaggataccgagggcctgaaggaagtcattaacgaccgctat
B.a.- PA        2008 ..t.....t..a..a..t..a.....t..a..g..t..a.....t..a..t...a.a...

TPA-human fu    2044 gatatgctgaacatttccagcttacgacaggacggtaagacatttattgactttaaaaag
B.a.- PA        2068 ......t....t.....t..t.....g..a..t..a..a........a..t........a

TPA-human fu    2104 tataacgacaagctacccctgtacatttccaacccaaattacaaagttaatgtgtatgct
B.a.- PA        2128 .....t..t..at....gt.a..t..aagt..t..c.....t..g..a.....a......

TPA-human fu    2164 gtaaccaaggagaacacaatcatcaatccaagcgagaacggcgataccagcacaaatgga
B.a.- PA        2188 ..t..t..a..a.....t..t..t.....t..t.....t..g.....t..t..c..c..g

TPA-human fu    2224 atcaaaaagatccttatatttagtaaaaaaggctacgagatcggttgaggatcc
B.a.- PA        2248 .....g..a..tt.a..c...tc...........t.....a..a.a.------
```

Figure 11A

```
TPA-Human LF     82 gccggcgggcatggggacgttggcatgcatgtgaaagaaaaggagaaaaacaaggacgaa
B.a.- LF        100 ..g.....t.....t..t..a..t.....c..a.....g..a........t..a..t..g

TPA-Human LF    142 aacaagcgtaaagacgaagaacgtaataaaacacaggaggaacacttaaaggagatcatg
B.a.- LF        160 ..t...a.a.....t........a..............a..g..t........a......

TPA-Human LF    202 aagcacatagtaaagattgaggtaaaaggcgaagaggctgtaaagaaggaggcagcagaa
B.a.- LF        220 ..a.....t.....a..a..a........g..g..a.....t..a..a............

TPA-Human LF    262 aaactgttggagaaggtgccttctgacgtcttagagatgtataaggccatcggcggtaag
B.a.- LF        280 ..g..ac.t.....a..a..a.....t..t..............a..a..t..a..a...

TPA-Human LF    322 atctatatcgtggacggagacatcactaaacacatatctctcgaagctctctccgaggac
B.a.- LF        340 ..a.....t.....t..t..t..t..a.....t......t.a.....at.a..t..a..t

TPA-Human LF    382 aagaaaaagattaaagacatctacgggaaggatgccttattgcacgagcactacgtttac
B.a.- LF        400 ........a..a........t..t.....a.....t.....a..t..a..t..t..a..t

TPA-Human LF    442 gcaaaggagggctatgagcccgtgctcgttattcagagtagtgaggactacgtcgagaat
B.a.- LF        460 .....a..a..a.....a.....a..t..a..c..atc.tcg..a..t..t..a..a...

TPA-Human LF    502 accgagaaagctctgaatgtgtattacgagatcggaaagattctgtcccgggacatcctg
B.a.- LF        520 ..t..a..g..a.....c..t.....t..a..a..t.....at.a..aa....t..tt.a

TPA-Human LF    562 tccaaaatcaaccagccataccagaaattccttgatgttcttaacacaatcaaaaacgcg
B.a.- LF        580 agt.....t..t..a.....t........tt.a.....at.a..t..c..t.....t..a

TPA-Human LF    622 tcagatagcgacgggcaggatcttctgtttacaaatcaactcaaggaacaccccactgat
B.a.- LF        640 ..t..tca...t..a..a.....t..a.....t.....g..t........t.....a..c

TPA-Human LF    682 ttcagcgtggagttcctcgagcagaattctaacgaagtccaggaggtgttcgccaaggca
B.a.- LF        700 ..ttct..a..a...t.g..a..a...agc..t..g..a..a..a..a..t..g..a..t

TPA-Human LF    742 tttgcgtactatatcgaaccccagcatcgcgatgtgctccagctgtacgccccggaggca
B.a.- LF        760 .....a..t........g..a........t.....tt.a.....t..t..a.....a..t

TPA-Human LF    802 tttaactacatggacaaattcaatgaacaggagattaatctgtctctggaggaactgaaa
B.a.- LF        820 .....t........t.....t..c.....a..a..a.....a..ct....a.....t...

TPA-Human LF    862 gaccagaggatgctctcccggtatgaaaagtgggaaaagatcaaacagcattaccagcat
B.a.- LF        880 ..t..ac.......g..aa.a........a...........a........c..t..a..c

TPA-Human LF    922 tggtccgactccctgtcagaagagggcgcggcctgttgaaaaagttgcagattcccatc
B.a.- LF        940 ...ag...t..tt.a..t.....a..aa.a..a..t..a......c..........t..t

TPA-Human LF    982 gagcctaagaaagatgatataatacactctctaagccaggaggagaaggaactcctgaag
B.a.- LF       1000 .....a...........c.....t..t...t..tct..a..a..a..a..g..t..a..a

TPA-Human LF   1042 cggatacaaatcgactcatccgatttccttagcacagaagagaaggagtttctaaaaaa
B.a.- LF       1060 a.a........t..tagtagt.....tt.atct..t..g..a..a......t.......g

TPA-Human LF   1102 cttcagatagatattagagattcactgagcgaggaagagaaggagctgctcaaccgaatt
B.a.- LF       1120 ..a..a..t......c.t.....tt.atct..a.....a..a.....tt.a..ta....a

TPA-Human LF   1162 caagtcgatagttcgaaccccttgtcagaaaaagagaaggaattcctgaaaaagttgaag
B.a.- LF       1180 ..g..g......agt..t..t..a..t........a..a..g..tt.a.....c.....a

TPA-Human LF   1222 ctcgacatccagccgtacgatattaatcagcggctacaagacaccggcggtctgattgat
B.a.- LF       1240 ..t..t..t..a..a..t...........aa..t.g.....t..a..a..gt.a......

TPA-Human LF   1282 agccccagcatcaaccttgacgtacggaagcaatataagcgcgacattcaaatatcgac
B.a.- LF       1300 ..t..gtca..t..t.....t...a.a.....g.....aa.g..t...........t..t

TPA-Human LF   1342 gccctattacatcaatccataggctccacgctatacaataaaatctatctatacgaaaac
B.a.- LF       1360 ..tt...............t..aagt..ct.g...........t...t.g..t.....t
```

Figure 11B

```
TPA-Human LF   1402 atgaatattaacaatctcaccgctacactgggagcggacctggtcgatagtacagacaac
B.a.- LF       1420 ........c..t..c..t..a..a..c..a..t.....tt.a..t...tcc..t..t..t

TPA-Human LF   1462 acaaagataaacagaggtattttcaacgaattcaaaaagaactttaagtattcgatcagc
B.a.- LF       1480 ..t..a..t..t...............t..........a..t..c..a...agt..ttct

TPA-Human LF   1522 agtaactatatgattgttgacatcaatgaacggcccgcattagacaatgagaggttgaag
B.a.- LF       1540 ..................t..a......a.....t........t.......c.t.....a

TPA-Human LF   1582 tggagaattcaactgagtcctgatactagggccggctatctggagaacgggaaactgatc
B.a.- LF       1600 ........c...t.atca..a......c.a..a..a...t.a..a..t..a..g..t..a

TPA-Human LF   1642 ttacagcgaaacatcgggctggagatcaaggatgtgcagattatcaagcagagcgaaaaa
B.a.- LF       1660 .....aa.........t.....a..a........a..a..a..t.....atc.......

TPA-Human LF   1702 gaatacattcgcatcgacgccaaggtggtgcctaagtcaaagatcgataccaagatccag
B.a.- LF       1720 .....t..aa.g..t..t..g..a..a......a...agt..a..a......a..a..t..a

TPA-Human LF   1762 gaagctcagctcaacattaaccaggagtggaataaagctcttggtctgccaaaatacacc
B.a.- LF       1780 .....a...t.a..t..a..t.....a...........at.a..gt.a.........t..a

TPA-Human LF   1822 aaacttatcacctttaatgtgcacaacaggtatgcctctaatatcgtcgagtcagcatac
B.a.- LF       1840 ..g.....t..a..c..c.....t..t..a.....a..c.....t..a..aagt..t..t

TPA-Human LF   1882 ctgattctcaatgaatggaagaacaatattcagtctgacctgatcaagaaggtcacgaat
B.a.- LF       1900 t.a..at.g...........a..t........aag...t..t..a..a.....a..a...

TPA-Human LF   1942 tatctggtggacggaaatggcagattcgtgttcaccgacataactttgccaaacattgcc
B.a.- LF       1960 ..ct.a..t..t..t.....a.....t..t..t.....t..t...c.c..t..t..a..t

TPA-Human LF   2002 gagcaatacactcatcaggatgaaatttacgagcaagtccactccaaaggtctgtatgtt
B.a.- LF       2020 ..a.....t..a.....a.....g..a..t........t..t..a.....gt.a......

TPA-Human LF   2062 ccagagtcaagatcgattctgctccatggtccatccaaaggggttgagcttcgaaacgat
B.a.- LF       2080 .....a..cc.t..t..at.a........a..t..a.....t..a..at.aa.g..t...

TPA-Human LF   2122 tctgagggatttatcgctgactttggagccgctgtggatgactacgccggatacctgttg
B.a.- LF       2140 ag......t.....acac..a......cat..........t..t..t.....t..a..a

TPA-Human LF   2182 gataagaatcagtctgatctcgtgacaaatagcaaaaaattcatagatattttcaaggag
B.a.- LF       2200 ........c..a......t.a..t......tct..........t........t.....a

TPA-Human LF   2242 gaagggagtaacctgacttcctatggccgcacgaacgaggctgaatttttgcggaagcc
B.a.- LF       2260 ...........tt.a.....g.....ga.a..a..t..a..g..........a......

TPA-Human LF   2302 tttagacttatgcacagcaccgaccatgctgaaaggttgaaggtgcaaaagaatgcccct
B.a.- LF       2320 .....gt.a.....ttct..g............c.t..a..a..t.....a.....t..g

TPA-Human LF   2362 aaaaccttccagttcataaatgaccagatcaagttcatcatcaactcttgaggatcc
B.a.- LF       2380 .....t.....a..t..t..c..t.....t........t..t.....a.a.------
```

Figure 12

```
TPA-Human PA     1 mdamkrglccvlllcgavfvspssagptvpdrdndgipdslevegytvdvknkrtflspw
Sugar minus      1 ............................................................

TPA-Human PA    61 isnihekkgltkykssspekwstasdpysdfekvtgridknvspearhplvaaypivhvdm
Sugar minus     61 .......................................q....................

TPA-Human PA   121 eniilsknedqstqntdsetrtisknts tsrthtsevhgnaevhasffdiggsvsagfsn
Sugar minus    121 .........................q..................................

TPA-Human PA   181 snsstvaidhslslagertwaetmglntadtarlnaniryvntgtapiynvlpttslvlg
Sugar minus    181 .q..........................................................

TPA-Human PA   241 knqtlatikakenqlsqilapnnyypsknlapialnaqddfsstpitmnynqflelektk
Sugar minus    241 .q..........................................................

TPA-Human PA   301 qlrldtdqvygniatynfengrvrvdtgsnwsevlpqiqettariifngkdlnlverria
Sugar minus    301 .............................q..............................

TPA-Human PA·  361 avnpsdplettkpdmtlkealkiafgfnepngnlqyqgkditefdfnfdqqtsqniknql
Sugar minus    361 ..q.........................................................

TPA-Human PA   421 aelnatniytvldkiklnakmnilirdkrfhydrnniavgadesvvkeahrevinssteg
Sugar minus    421 ...q..................................................q.....

TPA-Human PA   481 lllnidkdirkilsgyiveiedteglkevindrydmlnisslrqdgktfidfkkyndklp
Sugar minus    481 .....................................q......................

TPA-Human PA   541 lyisnpnykvnvyavtkentiinpsengdtstngikkilifskkgyeig
Sugar minus    541 ......................q..........................
```

Figure 13

```
TPA-Human LF   1 mdamkrglccvlllcgavfvspsagghgdvgmhvkekeknkdenkrkdeernktqeehlk
Sugar minus    1 ...................................................q........

TPA-Human LF  61 eimkhivkievkgeeavkkeaaeekllekvpsdvlemykaiggkiyivdgditkhisleal
Sugar minus   61 ............................................................

TPA-Human LF 121 sedkkkikdiygkdallhehyvyakegyepvlviqssedyventekalnvyyeigkilsr
Sugar minus  121 ............................................................

TPA-Human LF 181 dilskinqpyqkfldvlntiknasdsdgqdllftnqlkehptdfsvefleqnsnevqevf
Sugar minus  181 .....................q......................................

TPA-Human LF 241 akafayyiepqhrdvlqlyapeafnymdkfneqeinlsleelkdqrmlsryekwekikqh
Sugar minus  241 ...................................q........................

TPA-Human LF 301 yqhwsdslseegrgllkklqipiepkkddiihslsqeekellkriqidssdflsteekef
Sugar minus  301 ............................................................

TPA-Human LF 361 lkklqidirdslseeekellnriqvdssnplsekekeflkklkldiqpydinqrlqdtgg
Sugar minus  361 ............................................................

TPA-Human LF 421 lidspsinldvrkqykrdiqnidallhqsigstlynkiylyenmninnltatlgadlvds
Sugar minus  421 ...............................................q............

TPA-Human LF 481 tdntkinrgifnefkknfkysissnymivdinerpaldnerlkwriqlspdtragyleng
Sugar minus  481 ............................................................

TPA-Human LF 541 klilqrnigleikdvqiikqsekeyiridakvvpkskidtkiqeaqlninqewnkalglp
Sugar minus  541 ............................................................

TPA-Human LF 601 kytklitfnvhnryasnivesaylilnewknniqsdlikkvtnylvdgngrfvftditlp
Sugar minus  601 ............................................................

TPA-Human LF 661 niaeqythqdeiyeqvhskglyvpesrsillhgpskgvelrndsegfiadfgaavddyag
Sugar minus  661 .........................................q..................

TPA-Human LF 721 ylldknqsdlvtnskkfidifkeegsnltsygrtneaeffaeafrlmhstdhaerlkvqk
Sugar minus  721 .....q....................q.................................

TPA-Human LF 781 napktfqfindqikfiins
Sugar minus  781 ...................
```

Figure 14

```
  1  gatatcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga
                    m  d   a  m  k  r   g  l  c   c  v  l   l  l  c  g

 61  gcagtcttcg tttcgcccag cgccggcggg catggggacg ttggcatgca tgtgaaagaa
      a  v  f   v  s  p   s  a  g  g   h  g  d   v  g  m   h  v  k  e

121  aaggagaaaa acaaggacga aaacaagcgt aaagacgaag aacgtaataa aacacaggag
      k  e  k   n  k  d   e  n  k  r   k  d  e   e  r  n   k  t  q  e

181  gaacacttaa aggagatcat gaagcacata gtaaagattg aggtaaaagg cgaagaggct
      e  h  l   k  e  i   m  k  h  i   v  k  i   e  v  k   g  e  e  a

241  gtaaagaagg aggcagcaga aaaactgttg gagaaggtgc cttctgacgt cttagagatg
      v  k  k   e  a  a   e  k  l  l   e  k  v   p  s  d   v  l  e  m

301  tataaggcca tcggcggtaa gatctatatc gtggacggag acatcactaa acacatatct
      y  k  a   i  g  g   k  i  y  i   v  d  g   d  i  t   k  h  i  s

361  ctcgaagctc tctccgagga caagaaaaag attaaagaca tctacgggaa ggatgcctta
      l  e  a   l  s  e   d  k  k  k   i  k  d   i  y  g   k  d  a  l

421  ttgcacgagc actacgttta cgcaaaggag ggctatgagc ccgtgctcgt tattcagagt
      l  h  e   h  y  v   y  a  k  e   g  y  e   p  v  l   v  i  q  s

481  agtgaggact acgtcgagaa taccgagaaa gctctgaatg tgtattacga gatcggaaag
      s  e  d   y  v  e   n  t  e  k   a  l  n   v  y  y   e  i  g  k

541  attctgtccc gggacatcct gtccaaaatc aaccagccat accagaaatt ccttgatgtt
      i  l  s   r  d  i   l  s  k  i   n  q  p   y  q  k   f  l  d  v

601  cttaacacaa tcaaaaacgc gtcagatagc gacgggcagg atcttctgtt tacaaatcaa
      l  n  t   i  k  n   a  s  d  s   d  g  q   d  l  l   f  t  n  q

661  ctcaaggaac accccactga tttcagcgtg gagttcctcg agcagaattc taacgaagtc
      l  k  e   h  p  t   d  f  s  v   e  f  l   e  q  n   s  n  e  v

721  caggaggtgt tcgccaaggc atttgcgtac tatatcgaac cccagcatcg cgatgtgctc
      q  e  v   f  a  k   a  f  a  y   y  i  e   p  q  h   r  d  v  l

781  cagctgtacg ccccggaggc atttaactac atggacaaat tcaatgaaca ggagattaat
      q  l  y   a  p  e   a  f  n  y   m  d  k   f  n  e   q  e  i  n

841  ctgtctctgg aggaactgaa agaccagtga ggatcc
      l  s  l   e  e  l   k  d  q  -
```

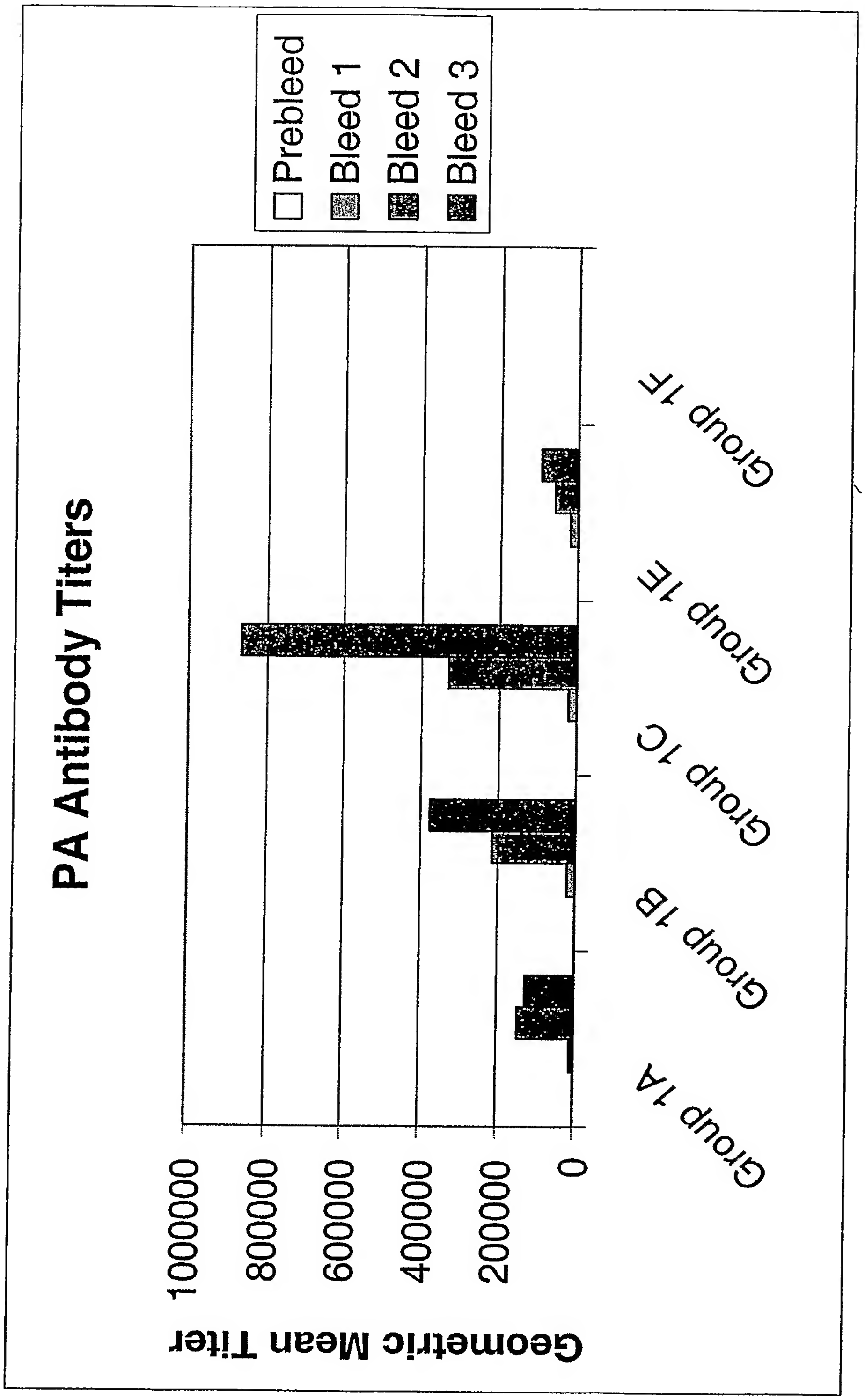
PA Antibody Titers
Prebleed
Bleed 1
Bleed 2
Bleed 3
Geometric Mean Titer
1000000
800000
600000
400000
200000
0
Group 1A
Group 1B
Group 1C
Group 1E
Group 1F

FIG. 15A

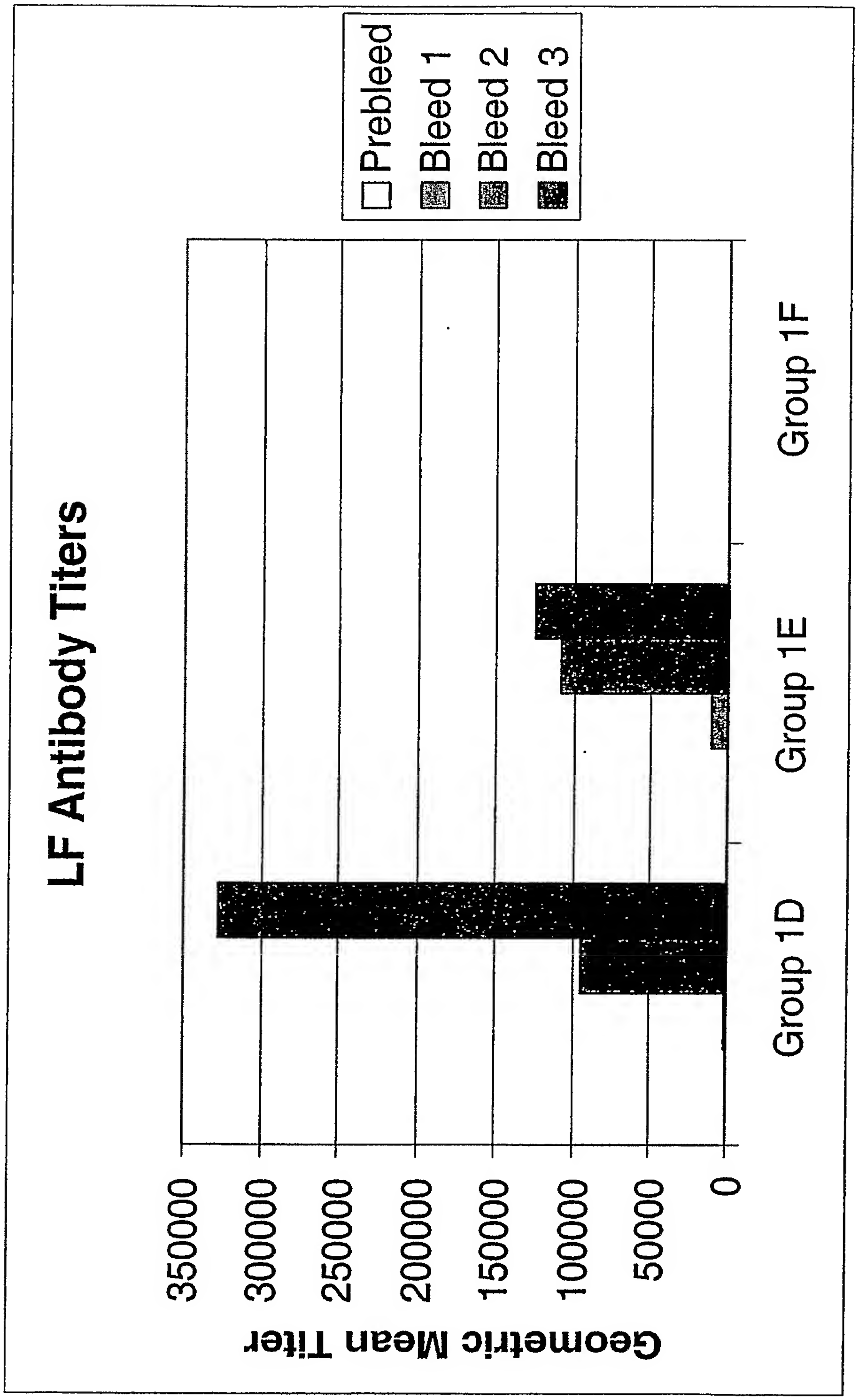
LF Antibody Titers
Prebleed
Bleed 1
Bleed 2
Bleed 3
Geometric Mean Titer
350000
300000
250000
200000
150000
100000
50000
0
Group 1D
Group 1E
Group 1F

FIG. 15B

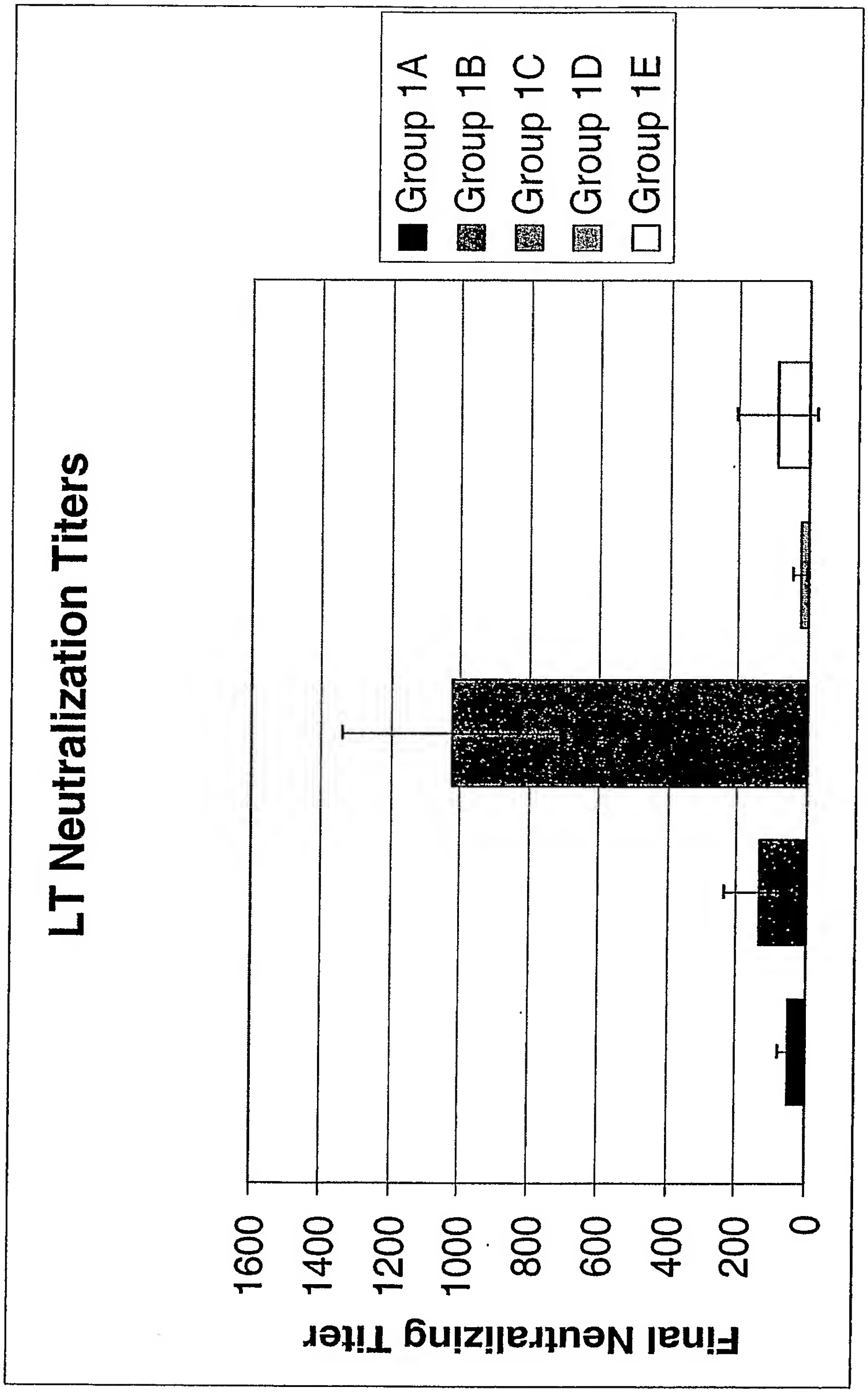
LT Neutralization Titers
Final Neutralizing Titer
1600
1400
1200
1000
800
600
400
200
0
Group 1A
Group 1B
Group 1C
Group 1D
Group 1E

FIG. 15C

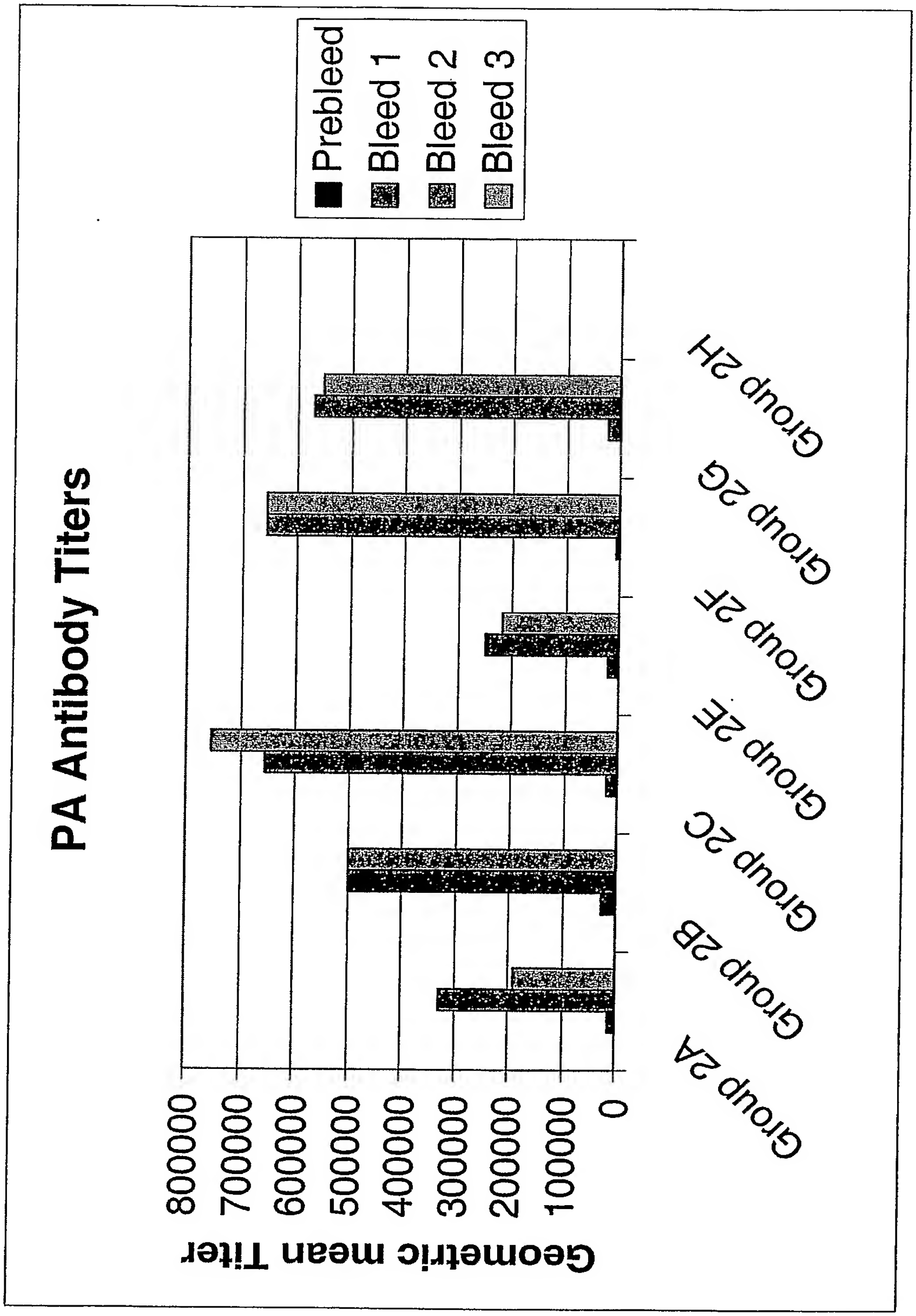
PA Antibody Titers
Geometric mean Titer
800000
700000
600000
500000
400000
300000
200000
100000
0
Group 2A
Group 2B
Group 2C
Group 2E
Group 2F
Group 2G
Group 2H
Prebleed
Bleed 1
Bleed 2
Bleed 3

FIG. 16A

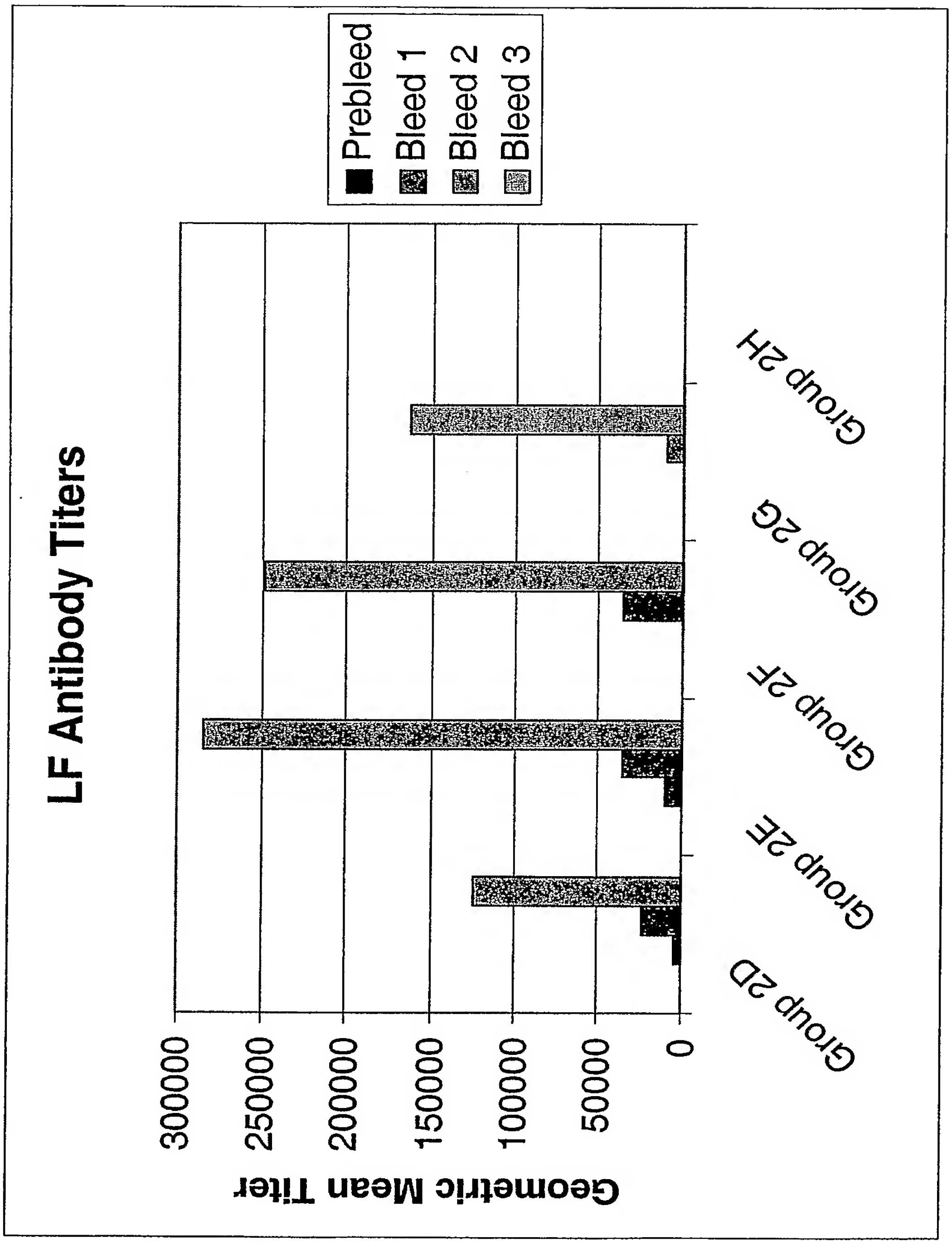
LF Antibody Titers
Geometric Mean Titer
300000
250000
200000
150000
100000
50000
0
Group 2D
Group 2E
Group 2F
Group 2G
Group 2H
Prebleed
Bleed 1
Bleed 2
Bleed 3

FIG. 16B

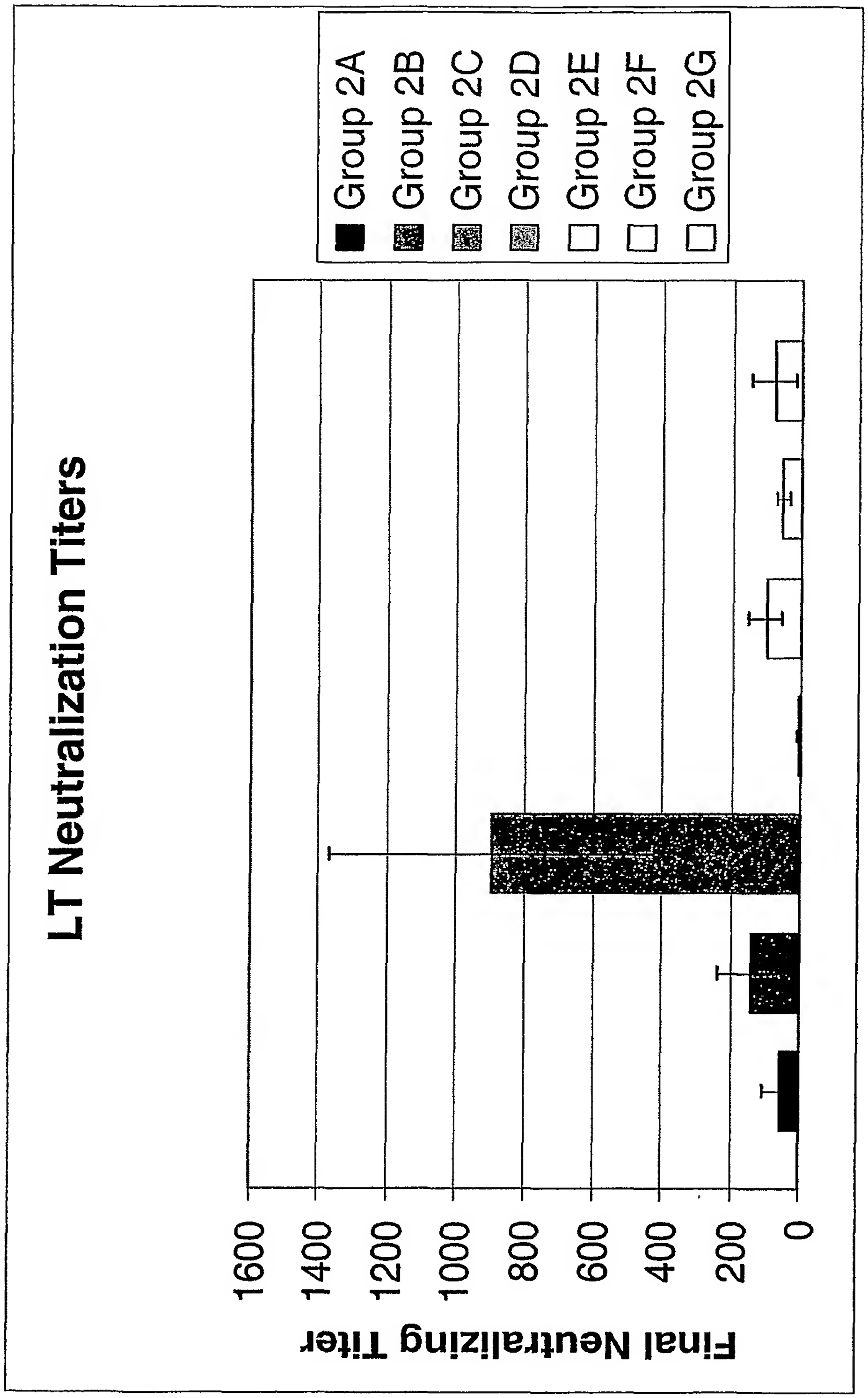
LT Neutralization Titers
Final Neutralizing Titer
1600
1400
1200
1000
800
600
400
200
0
Group 2A
Group 2B
Group 2C
Group 2D
Group 2E
Group 2F
Group 2G

FIG. 16C

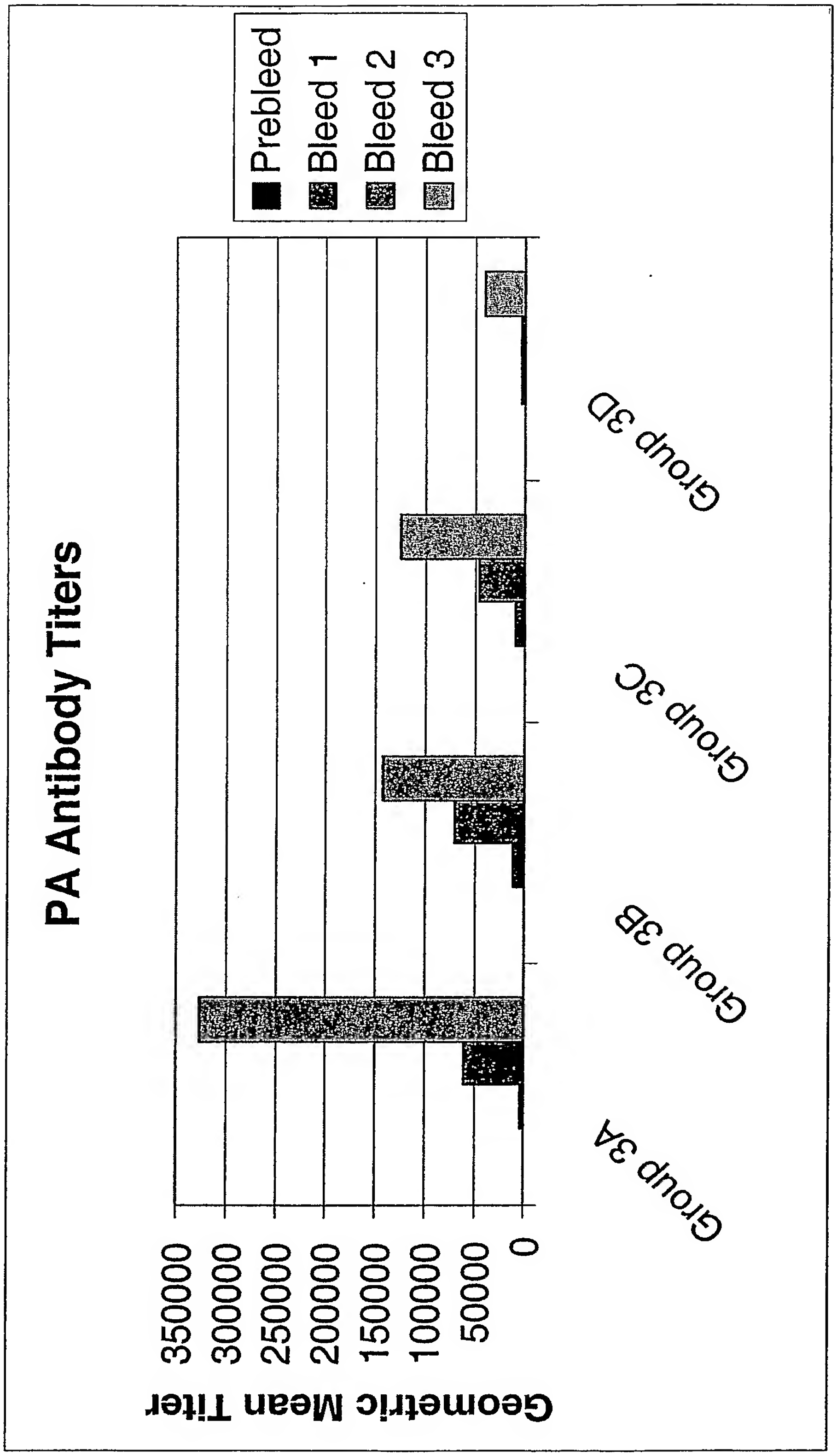
PA Antibody Titers
Geometric Mean Titer
350000
300000
250000
200000
150000
100000
50000
0
Group 3A
Group 3B
Group 3C
Group 3D
Prebleed
Bleed 1
Bleed 2
Bleed 3

FIG. 17A

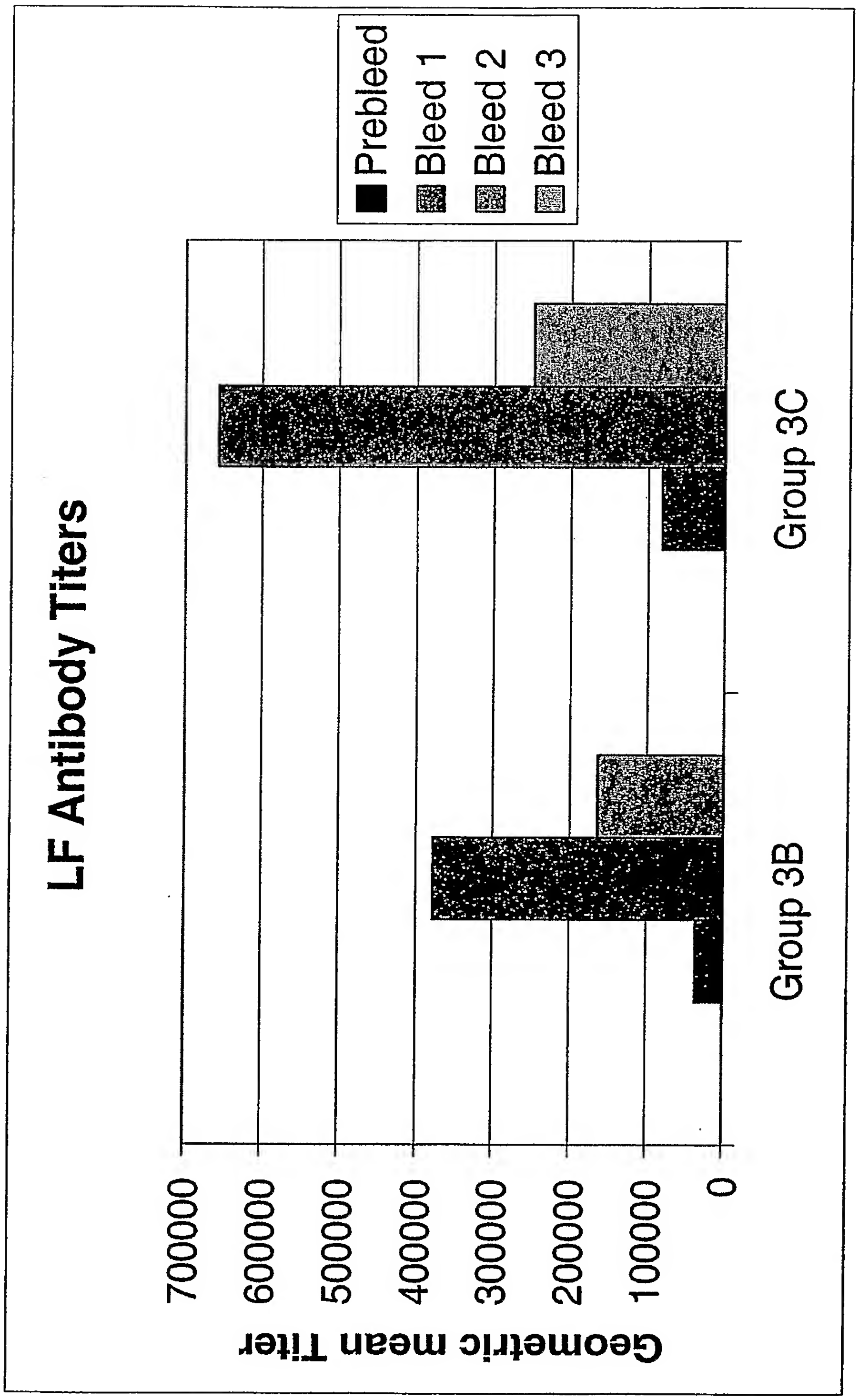
LF Antibody Titers
Geometric mean Titer
700000
600000
500000
400000
300000
200000
100000
0
Group 3B
Group 3C
Prebleed
Bleed 1
Bleed 2
Bleed 3

FIG. 17B

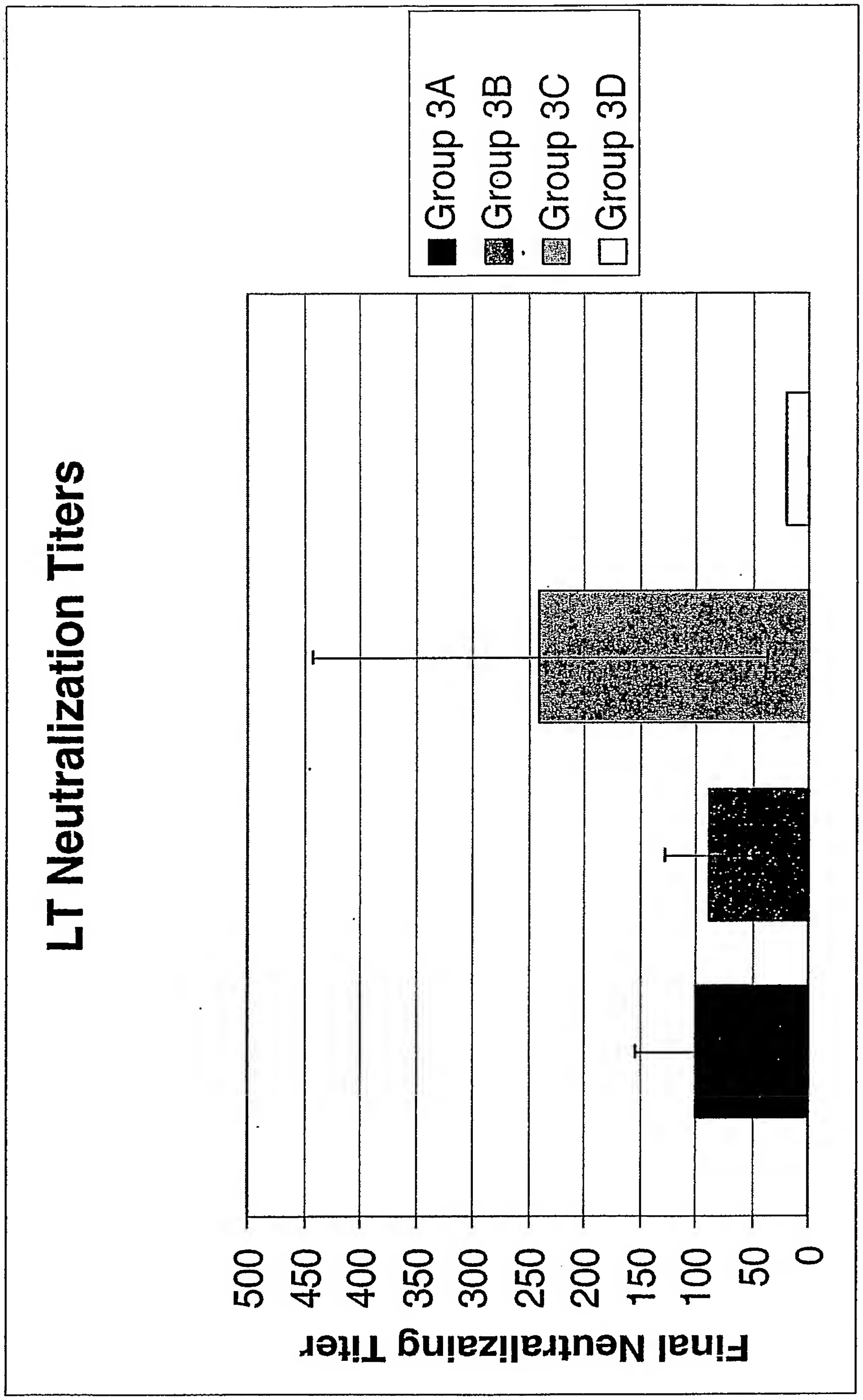
LT Neutralization Titers
Final Neutralizing Titer
500
450
400
350
300
250
200
150
100
50
0
Group 3A
Group 3B
Group 3C
Group 3D

FIG. 17C

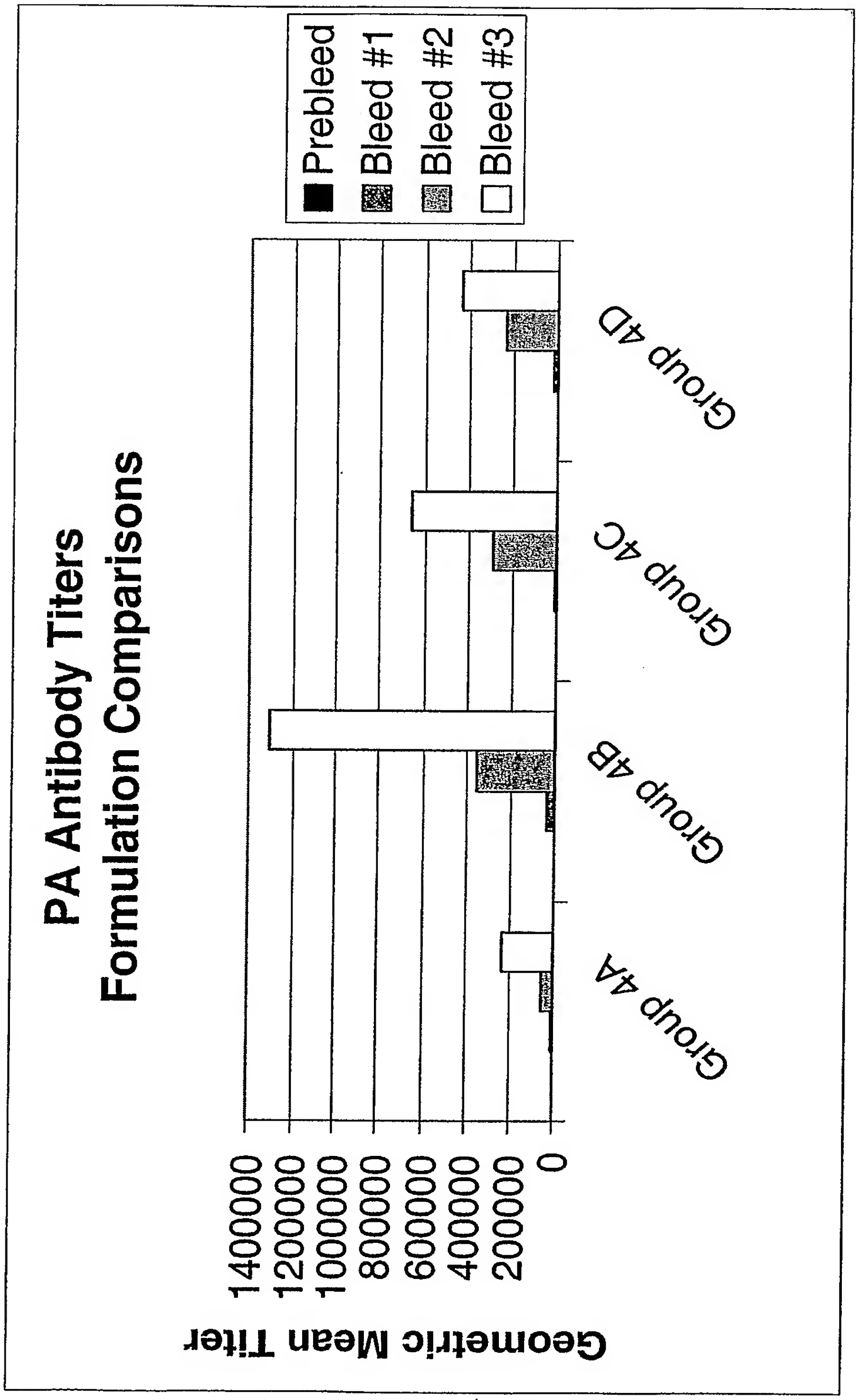
PA Antibody Titers
Formulation Comparisons
Geometric Mean Titer
1400000
1200000
1000000
800000
600000
400000
200000
0
Group 4A
Group 4B
Group 4C
Group 4D
Prebleed
Bleed #1
Bleed #2
Bleed #3

FIG. 18A

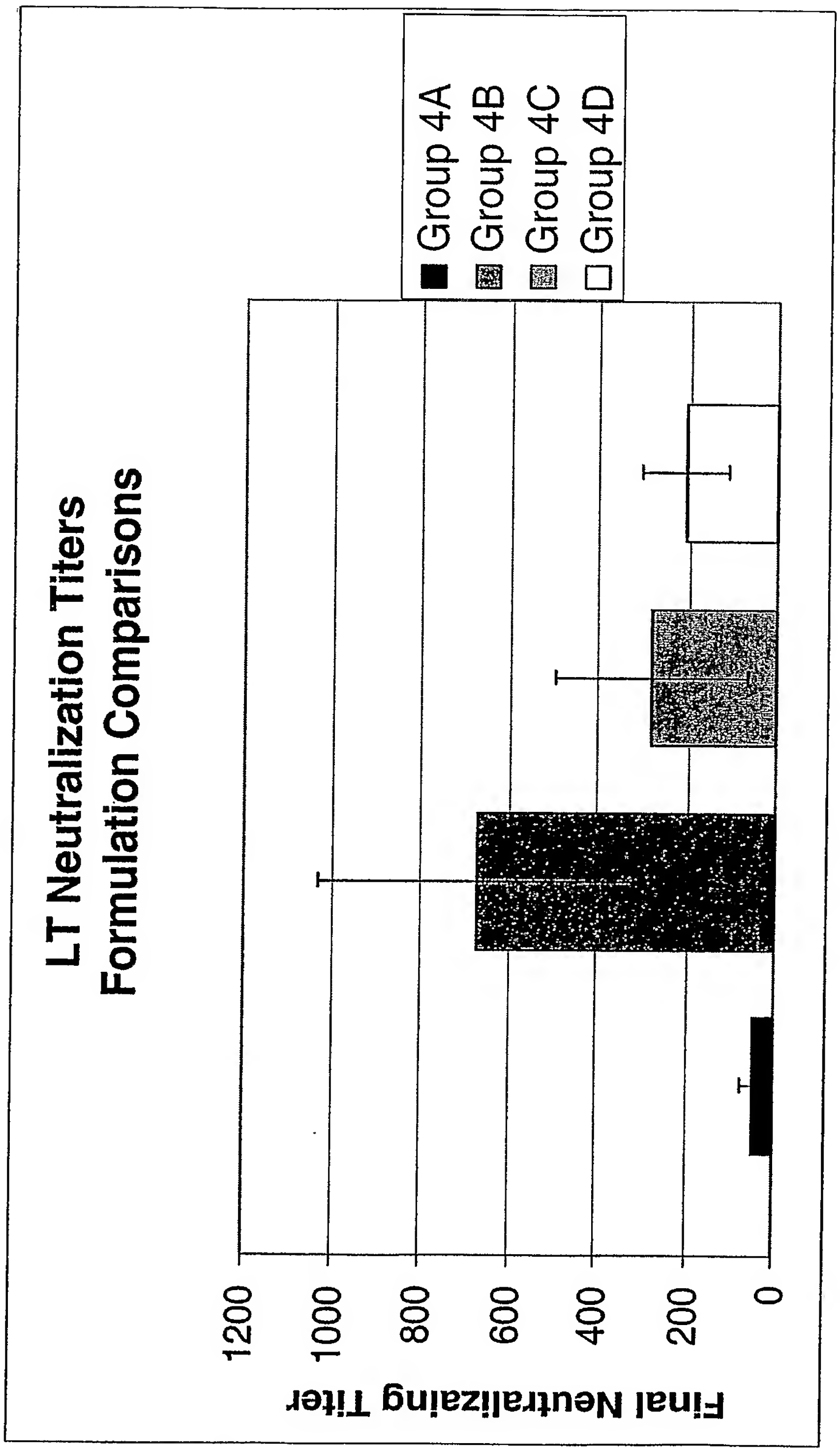
LT Neutralization Titers
Formulation Comparisons
Final Neutralizing Titer
1200
1000
800
600
400
200
0
Group 4A
Group 4B
Group 4C
Group 4D

FIG. 18B

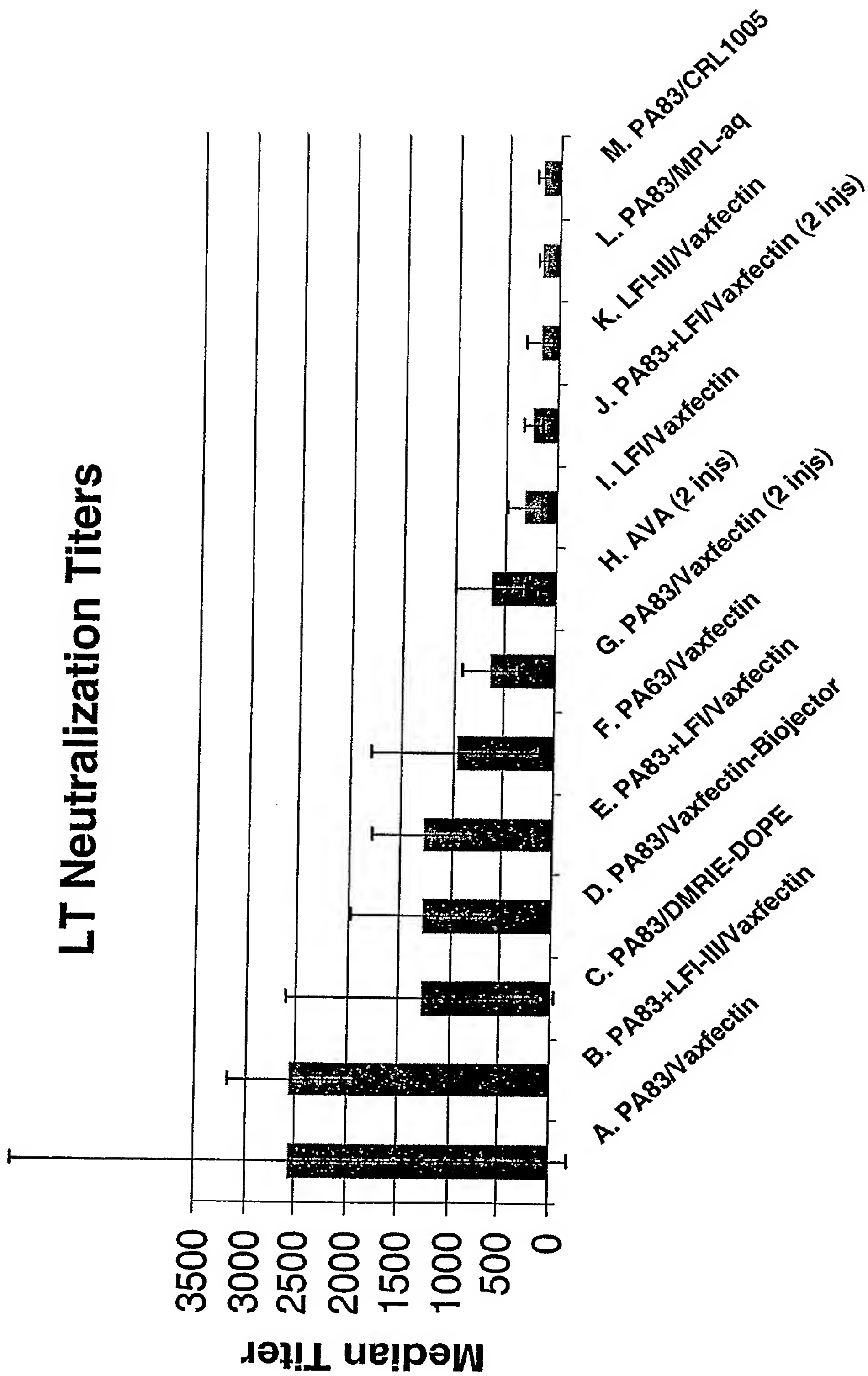
LT Neutralization Titers
Median Titer
3500
3000
2500
2000
1500
1000
500
0
A. PA83/Vaxfectin
B. PA83+LFI-III/Vaxfectin
C. PA83/DMRIE-DOPE
D. PA83/Vaxfectin-Biojector
E. PA83+LFI/Vaxfectin
F. PA63/Vaxfectin
G. PA83/Vaxfectin (2 injs)
H. AVA (2 injs)
I. LFI/Vaxfectin
J. PA83+LFI/Vaxfectin (2 injs)
K. LFI-III/Vaxfectin
L. PA83/MPL-aq
M. PA83/CRL1005

FIG. 19

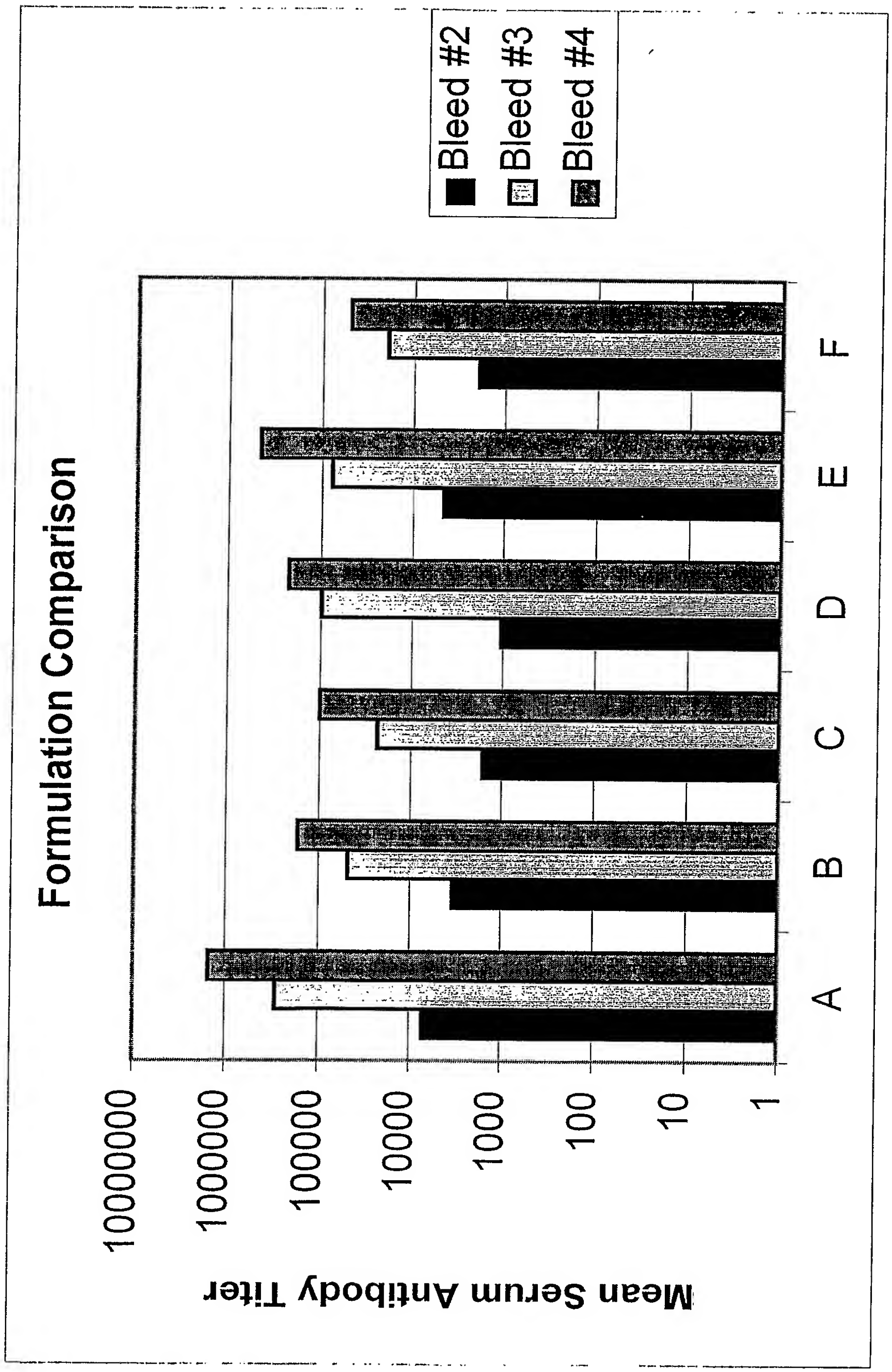
Formulation Comparison
Mean Serum Antibody Titer
10000000
1000000
100000
10000
1000
100
10
1
A
B
C
D
E
F
Bleed #2
Bleed #3
Bleed #4

FIGURE 20

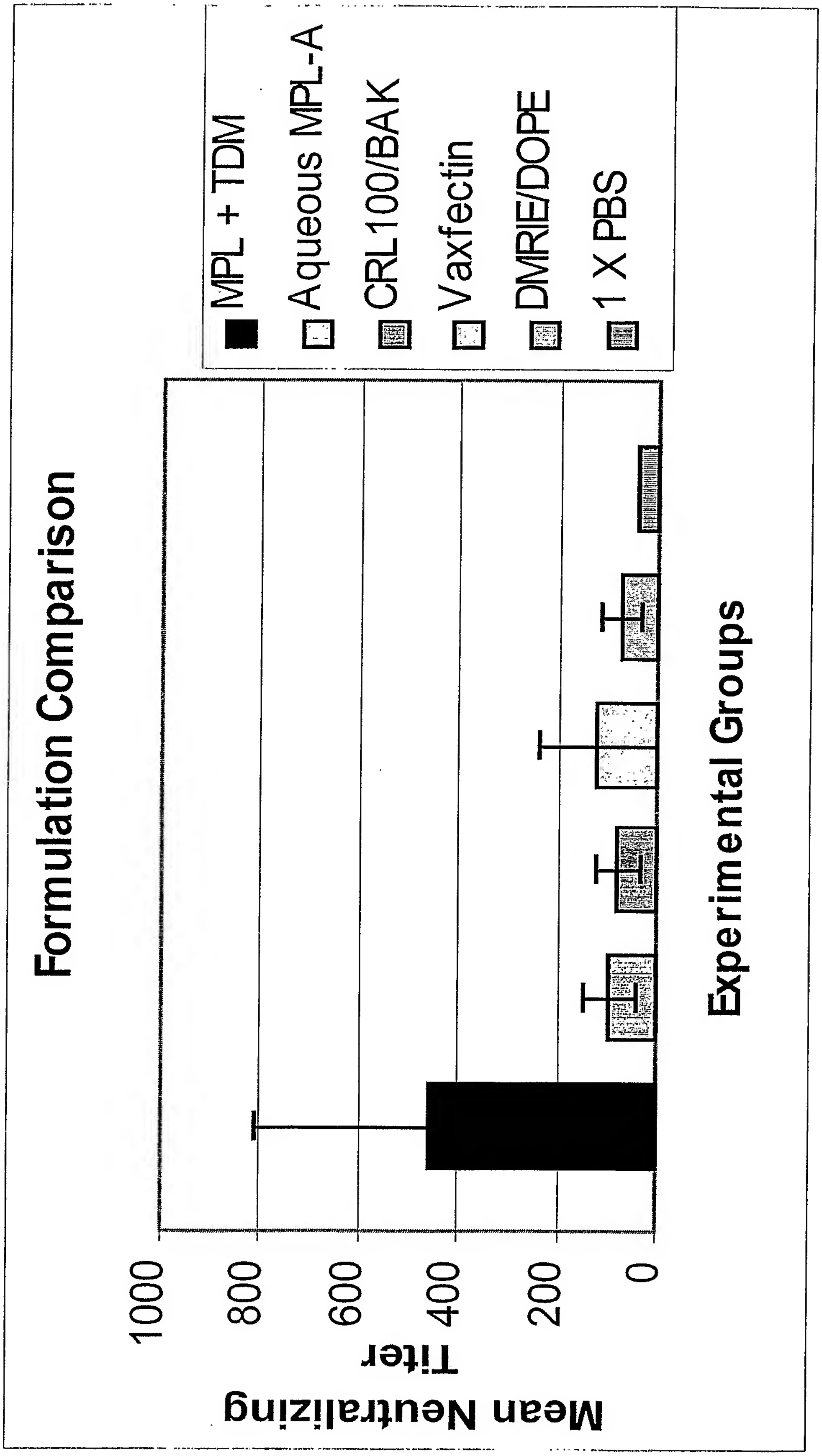
Formulation Comparison
Mean Neutralizing Titer
1000
800
600
400
200
0
Experimental Groups
MPL + TDM
Aqueous MPL-A
CRL100/BAK
Vaxfectin
DMRIE/DOPE
1 X PBS

FIGURE 21

SEQUENCE LISTING

<110> Vical Incorporated

<120> Codon-Optimized Polynucleotide-Based Vaccines
Against Bacillus anthracis Infection

<130> 1530.046PC03

<150> US 60/409,307
<151> 2002-09-10

<150> US 60/419,089
<151> 2002-10-18

<160> 76

<170> PatentIn version 3.1

<210> 1

<211> 1788

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic coding region for Human TPA/B. anthracis
antigen fusion protein

<220>

<221> CDS

<222> (13)..(1779)

<223>

<400> 1
```
gatatcgcca cc atg gat gca atg aag aga ggg ctc tgc tgt gtg ctg ctg      51
              Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu
              1               5                   10

ctg tgt gga gca gtc ttc gtt tcg ccc agc agc gct ggg cca act gtg        99
```

```
Leu Cys Gly Ala Val Phe Val Ser Pro Ser Ser Ala Gly Pro Thr Val
    15                  20                  25

ccc gac aga gac aat gat gga atc cct gat agt cta gag gtt gag gga      147
Pro Asp Arg Asp Asn Asp Gly Ile Pro Asp Ser Leu Glu Val Glu Gly
30                  35                  40                  45

tac acg gta gat gtc aag aac aaa agg act ttt ctc tcg cct tgg atc      195
Tyr Thr Val Asp Val Lys Asn Lys Arg Thr Phe Leu Ser Pro Trp Ile
                50                  55                  60

tca aat atc cat gag aag aag ggg ctt acc aag tac aag tcc tcc ccc      243
Ser Asn Ile His Glu Lys Lys Gly Leu Thr Lys Tyr Lys Ser Ser Pro
            65                  70                  75

gag aag tgg tct acc gct tcc gat cca tat agc gat ttc gag aag gtc      291
Glu Lys Trp Ser Thr Ala Ser Asp Pro Tyr Ser Asp Phe Glu Lys Val
        80                  85                  90

aca ggc cgg atc gat aaa aat gtg tct cca gag gct aga cac ccc ctg      339
Thr Gly Arg Ile Asp Lys Asn Val Ser Pro Glu Ala Arg His Pro Leu
    95                  100                 105

gta gca gcc tac ccg att gta cac gtg gac atg gag aac atc att cta      387
Val Ala Ala Tyr Pro Ile Val His Val Asp Met Glu Asn Ile Ile Leu
110                 115                 120                 125

agc aaa aac gag gac cag tcc aca caa aac act gac tcc gag acc cgc      435
Ser Lys Asn Glu Asp Gln Ser Thr Gln Asn Thr Asp Ser Glu Thr Arg
                130                 135                 140

acc ata tct aaa aac acc agt act tca agg acc cac acc tct gaa gtg      483
Thr Ile Ser Lys Asn Thr Ser Thr Ser Arg Thr His Thr Ser Glu Val
            145                 150                 155

cac ggc aat gcg gaa gtc cat gca tcg ttt ttc gat att ggt ggc tcc      531
His Gly Asn Ala Glu Val His Ala Ser Phe Phe Asp Ile Gly Gly Ser
        160                 165                 170

gtg tca gcc ggc ttt agc aat agc aac tcc tcg acg gtt gcc att gac      579
Val Ser Ala Gly Phe Ser Asn Ser Asn Ser Ser Thr Val Ala Ile Asp
    175                 180                 185

cac tca ctg tca tta gca ggt gag agg act tgg gct gaa act atg ggt      627
His Ser Leu Ser Leu Ala Gly Glu Arg Thr Trp Ala Glu Thr Met Gly
190                 195                 200                 205

ctg aat acc gcc gat acg gcc cgg ctc aac gca aat att cgg tac gtc      675
Leu Asn Thr Ala Asp Thr Ala Arg Leu Asn Ala Asn Ile Arg Tyr Val
                210                 215                 220

aac aca ggg act gct cct ata tat aac gtg ctg cct acg aca agt ctt      723
Asn Thr Gly Thr Ala Pro Ile Tyr Asn Val Leu Pro Thr Thr Ser Leu
            225                 230                 235

gtc ctg ggc aaa aat cag acc ctc gca acc att aag gca aag gaa aat      771
Val Leu Gly Lys Asn Gln Thr Leu Ala Thr Ile Lys Ala Lys Glu Asn
        240                 245                 250

cag ctg agc cag atc ctc gcc cct aac aac tat tat cca tcc aaa aat      819
Gln Leu Ser Gln Ile Leu Ala Pro Asn Asn Tyr Tyr Pro Ser Lys Asn
    255                 260                 265
```

```
tta gcc ccc ata gcc ctg aac gcc cag gac gac ttt tcc tct acc ccc       867
Leu Ala Pro Ile Ala Leu Asn Ala Gln Asp Asp Phe Ser Ser Thr Pro
270                 275                 280                 285

ata act atg aat tac aat cag ttc ctg gag ctg gaa aag acg aag cag       915
Ile Thr Met Asn Tyr Asn Gln Phe Leu Glu Leu Glu Lys Thr Lys Gln
                290                 295                 300

ctg aga cta gac acc gat cag gtg tat gga aac ata gcg aca tat aac       963
Leu Arg Leu Asp Thr Asp Gln Val Tyr Gly Asn Ile Ala Thr Tyr Asn
            305                 310                 315

ttt gag aac ggc cgc gtg cgc gtc gac act ggg tca aac tgg tct gaa      1011
Phe Glu Asn Gly Arg Val Arg Val Asp Thr Gly Ser Asn Trp Ser Glu
        320                 325                 330

gtt ctg ccg caa att caa gag aca acc gcc aga att atc ttt aat ggg      1059
Val Leu Pro Gln Ile Gln Glu Thr Thr Ala Arg Ile Ile Phe Asn Gly
    335                 340                 345

aag gac ttg aac ctt gtc gaa cgt aga att gcc gcc gtg aac ccc agt      1107
Lys Asp Leu Asn Leu Val Glu Arg Arg Ile Ala Ala Val Asn Pro Ser
350                 355                 360                 365

gat cca ctc gag acg act aaa ccg gat atg aca ctg aaa gag gct ctg      1155
Asp Pro Leu Glu Thr Thr Lys Pro Asp Met Thr Leu Lys Glu Ala Leu
                370                 375                 380

aag att gcc ttc gga ttc aac gaa cct aat ggc aat ttg cag tat cag      1203
Lys Ile Ala Phe Gly Phe Asn Glu Pro Asn Gly Asn Leu Gln Tyr Gln
            385                 390                 395

ggg aaa gac atc aca gag ttt gat ttc aat ttc gat cag cag act tcc      1251
Gly Lys Asp Ile Thr Glu Phe Asp Phe Asn Phe Asp Gln Gln Thr Ser
        400                 405                 410

caa aat atc aaa aat cag ttg gca gag ctg aat gcc acc aat atc tac      1299
Gln Asn Ile Lys Asn Gln Leu Ala Glu Leu Asn Ala Thr Asn Ile Tyr
    415                 420                 425

acg gtt ctc gat aaa atc aaa ctt aac gcc aag atg aac ata ttg att      1347
Thr Val Leu Asp Lys Ile Lys Leu Asn Ala Lys Met Asn Ile Leu Ile
430                 435                 440                 445

cga gac aaa cgc ttc cac tac gac cgc aac aat ata gcc gta ggc gct      1395
Arg Asp Lys Arg Phe His Tyr Asp Arg Asn Asn Ile Ala Val Gly Ala
                450                 455                 460

gat gag tct gtc gtc aag gag gct cat agg gaa gtt atc aac agc agt      1443
Asp Glu Ser Val Val Lys Glu Ala His Arg Glu Val Ile Asn Ser Ser
            465                 470                 475

act gaa ggg ctg tta ctt aat atc gac aag gac att cgg aag atc ctg      1491
Thr Glu Gly Leu Leu Leu Asn Ile Asp Lys Asp Ile Arg Lys Ile Leu
        480                 485                 490

tcc ggg tat atc gtg gag atc gag gat acc gag ggc ctg aag gaa gtc      1539
Ser Gly Tyr Ile Val Glu Ile Glu Asp Thr Glu Gly Leu Lys Glu Val
    495                 500                 505

att aac gac cgc tat gat atg ctg aac att tcc agc tta cga cag gac      1587
Ile Asn Asp Arg Tyr Asp Met Leu Asn Ile Ser Ser Leu Arg Gln Asp
```

```
510                 515                 520                 525

ggt aag aca ttt att gac ttt aaa aag tat aac gac aag cta ccc ctg      1635
Gly Lys Thr Phe Ile Asp Phe Lys Lys Tyr Asn Asp Lys Leu Pro Leu
                530                 535                 540

tac att tcc aac cca aat tac aaa gtt aat gtg tat gct gta acc aag      1683
Tyr Ile Ser Asn Pro Asn Tyr Lys Val Asn Val Tyr Ala Val Thr Lys
            545                 550                 555

gag aac aca atc atc aat cca agc gag aac ggc gat acc agc aca aat      1731
Glu Asn Thr Ile Ile Asn Pro Ser Glu Asn Gly Asp Thr Ser Thr Asn
        560                 565                 570

gga atc aaa aag atc ctt ata ttt agt aaa aaa ggc tac gag atc ggt      1779
Gly Ile Lys Lys Ile Leu Ile Phe Ser Lys Lys Gly Tyr Glu Ile Gly
    575                 580                 585

tgaggatcc                                                            1788


<210>  2

<211>  589

<212>  PRT

<213>  Artificial Sequence



<220>

<223>  Human TPA/ B. anthracis antigen fusion protein

<400>  2

Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
1               5                   10                  15


Ala Val Phe Val Ser Pro Ser Ser Ala Gly Pro Thr Val Pro Asp Arg
            20                  25                  30


Asp Asn Asp Gly Ile Pro Asp Ser Leu Glu Val Glu Gly Tyr Thr Val
        35                  40                  45


Asp Val Lys Asn Lys Arg Thr Phe Leu Ser Pro Trp Ile Ser Asn Ile
    50                  55                  60


His Glu Lys Lys Gly Leu Thr Lys Tyr Lys Ser Ser Pro Glu Lys Trp
65                  70                  75                  80


Ser Thr Ala Ser Asp Pro Tyr Ser Asp Phe Glu Lys Val Thr Gly Arg
                85                  90                  95


Ile Asp Lys Asn Val Ser Pro Glu Ala Arg His Pro Leu Val Ala Ala
```

```
            100                     105                     110

Tyr Pro Ile Val His Val Asp Met Glu Asn Ile Ile Leu Ser Lys Asn
        115                     120                     125

Glu Asp Gln Ser Thr Gln Asn Thr Asp Ser Glu Thr Arg Thr Ile Ser
    130                     135                     140

Lys Asn Thr Ser Thr Ser Arg Thr His Thr Ser Glu Val His Gly Asn
145                 150                     155                 160

Ala Glu Val His Ala Ser Phe Phe Asp Ile Gly Gly Ser Val Ser Ala
                165                     170                     175

Gly Phe Ser Asn Ser Asn Ser Ser Thr Val Ala Ile Asp His Ser Leu
            180                     185                     190

Ser Leu Ala Gly Glu Arg Thr Trp Ala Glu Thr Met Gly Leu Asn Thr
        195                     200                     205

Ala Asp Thr Ala Arg Leu Asn Ala Asn Ile Arg Tyr Val Asn Thr Gly
    210                     215                     220

Thr Ala Pro Ile Tyr Asn Val Leu Pro Thr Thr Ser Leu Val Leu Gly
225                 230                     235                 240

Lys Asn Gln Thr Leu Ala Thr Ile Lys Ala Lys Glu Asn Gln Leu Ser
                245                     250                     255

Gln Ile Leu Ala Pro Asn Asn Tyr Tyr Pro Ser Lys Asn Leu Ala Pro
            260                     265                     270

Ile Ala Leu Asn Ala Gln Asp Asp Phe Ser Ser Thr Pro Ile Thr Met
        275                     280                     285

Asn Tyr Asn Gln Phe Leu Glu Leu Glu Lys Thr Lys Gln Leu Arg Leu
    290                     295                     300

Asp Thr Asp Gln Val Tyr Gly Asn Ile Ala Thr Tyr Asn Phe Glu Asn
305                 310                     315                 320

Gly Arg Val Arg Val Asp Thr Gly Ser Asn Trp Ser Glu Val Leu Pro
                325                     330                     335

Gln Ile Gln Glu Thr Thr Ala Arg Ile Ile Phe Asn Gly Lys Asp Leu
            340                     345                     350
```

```
Asn Leu Val Glu Arg Arg Ile Ala Ala Val Asn Pro Ser Asp Pro Leu
        355                 360                 365

Glu Thr Thr Lys Pro Asp Met Thr Leu Lys Glu Ala Leu Lys Ile Ala
    370                 375                 380

Phe Gly Phe Asn Glu Pro Asn Gly Asn Leu Gln Tyr Gln Gly Lys Asp
385                 390                 395                 400

Ile Thr Glu Phe Asp Phe Asn Phe Asp Gln Gln Thr Ser Gln Asn Ile
                405                 410                 415

Lys Asn Gln Leu Ala Glu Leu Asn Ala Thr Asn Ile Tyr Thr Val Leu
            420                 425                 430

Asp Lys Ile Lys Leu Asn Ala Lys Met Asn Ile Leu Ile Arg Asp Lys
        435                 440                 445

Arg Phe His Tyr Asp Arg Asn Asn Ile Ala Val Gly Ala Asp Glu Ser
    450                 455                 460

Val Val Lys Glu Ala His Arg Glu Val Ile Asn Ser Ser Thr Glu Gly
465                 470                 475                 480

Leu Leu Leu Asn Ile Asp Lys Asp Ile Arg Lys Ile Leu Ser Gly Tyr
                485                 490                 495

Ile Val Glu Ile Glu Asp Thr Glu Gly Leu Lys Glu Val Ile Asn Asp
            500                 505                 510

Arg Tyr Asp Met Leu Asn Ile Ser Ser Leu Arg Gln Asp Gly Lys Thr
        515                 520                 525

Phe Ile Asp Phe Lys Lys Tyr Asn Asp Lys Leu Pro Leu Tyr Ile Ser
    530                 535                 540

Asn Pro Asn Tyr Lys Val Asn Val Tyr Ala Val Thr Lys Glu Asn Thr
545                 550                 555                 560

Ile Ile Asn Pro Ser Glu Asn Gly Asp Thr Ser Thr Asn Gly Ile Lys
                565                 570                 575

Lys Ile Leu Ile Phe Ser Lys Lys Gly Tyr Glu Ile Gly
            580                 585
```

<210> 3

<211> 4235

<212> DNA

<213> Bacillus anthracis

<220>

<221> CDS

<222> (1804)..(4098)

<223>

<400> 3

```
aagcttctgt cattcgtaaa tttcaaatag aacgtaaatt tagacttctc atcattaaaa      60
atgaaaaatc ttatcttttt gattctattg tatatttttta ttaaggtgtt taatagttag     120
aaaagacagt tgatgctatt actccagata aaatatagct aaccataaat ttattaaaga     180
aaccttgttg ttctaaataa tgattttgtg gattccggaa tagatactgg tgagttagct     240
ctaattttat agtgatttaa ctaacaattt ataaagcagc ataattcaaa ttttttaatt     300
gatttttcct gaagcatagt ataaaagagt caaggtcttc tagacttgac tcttggaatc     360
attaggaatt aacaatatat ataatgcgct agacagaatc aaattaaatg caaaaatgaa     420
tattttagta agagatccat atcattatga taataacggt aatattgtag gggttgatga     480
ttcatattta aaaaacgcat ataagcaaat acttaattgg tcaagcgatg gagtttcttt     540
aaatctagat gaagatgtaa atcaagcact atctggatat atgcttcaaa taaaaaaacc     600
ttcaaaccac ctaacaaaca gcccagttac aattacatta gcaggcaagg acagtggtgt     660
tggagaattg tatagagtat tatcagatgg agcaggattc ctggatttca ataagtttga     720
tgaaaattgg cgatcattag tagatcctgg tgatgatgtt tatgtgtatg ctgttactaa     780
agaagatttt aatgcagtta ctcgagatga aaatggtaat atagcgaata aattaaaaaa     840
caccttagtt ttatcgggta aaataaaaga aataaacata aaaactacaa atattaatat     900
atttgtagtt tttatgttta ttatatacct cctattttat attattagta gcacagtttt     960
tgcaaatcat gtaattgtat acttatctat gtagaggtat cacaacttat gaatagtgta    1020
ttttattgaa cgttggttag cttggacagt tgtatggata tgcatacttt ataacgtata    1080
aaatttcacg caccacaata aaactaattt aacaaaaaca aaaacacacc taagatcatt    1140
cagttctttt aataaggagc tgcccaccaa gctaaaccta aataatcttt gtttcacata    1200
aggttttttt ctaaatatac agtgtaagtt attgtgaatt taaccagtat atattaaaaa    1260
tgttttatgt taacaaatta aattgtaaaa cccctcttaa gcatagttaa gaggggtagg    1320
```

ttttaaattt tttgttgaaa ttagaaaaaa taataaaaaa acaaacctat tttctttcag 1380

gttgtttttg ggttacaaaa caaaaagaaa acatgtttca aggtacaata attatggttc 1440

tttagctttc tgtaaaacag ccttaatagt tggatttatg actattaaag ttagtataca 1500

gcatacacaa tctattgaag gatatttata atgcaattcc ctaaaaatag ttttgtataa 1560

ccagttcttt tatccgaact gatacacgta ttttagcata atttttaatg tatcttcaaa 1620

aacagcttct gtgtcctttt ctattaaaca tataaattct tttttatgtt atatatttat 1680

aaaagttctg tttaaaaagc caaaaataaa taattatctc tttttattta tattatattg 1740

aaactaaagt ttattaattt caatataata taaatttaat tttatacaaa aaggagaacg 1800

tat atg aaa aaa cga aaa gtg tta ata cca tta atg gca ttg tct acg 1848
Met Lys Lys Arg Lys Val Leu Ile Pro Leu Met Ala Leu Ser Thr
1 5 10 15

ata tta gtt tca agc aca ggt aat tta gag gtg att cag gca gaa gtt 1896
Ile Leu Val Ser Ser Thr Gly Asn Leu Glu Val Ile Gln Ala Glu Val
20 25 30

aaa cag gag aac cgg tta tta aat gaa tca gaa tca agt tcc cag ggg 1944
Lys Gln Glu Asn Arg Leu Leu Asn Glu Ser Glu Ser Ser Ser Gln Gly
35 40 45

tta cta gga tac tat ttt agt gat ttg aat ttt caa gca ccc atg gtg 1992
Leu Leu Gly Tyr Tyr Phe Ser Asp Leu Asn Phe Gln Ala Pro Met Val
50 55 60

gtt acc tct tct act aca ggg gat tta tct att cct agt tct gag tta 2040
Val Thr Ser Ser Thr Thr Gly Asp Leu Ser Ile Pro Ser Ser Glu Leu
65 70 75

gaa aat att cca tcg gaa aac caa tat ttt caa tct gct att tgg tca 2088
Glu Asn Ile Pro Ser Glu Asn Gln Tyr Phe Gln Ser Ala Ile Trp Ser
80 85 90 95

gga ttt atc aaa gtt aag aag agt gat gaa tat aca ttt gct act tcc 2136
Gly Phe Ile Lys Val Lys Lys Ser Asp Glu Tyr Thr Phe Ala Thr Ser
100 105 110

gct gat aat cat gta aca atg tgg gta gat gac caa gaa gtg att aat 2184
Ala Asp Asn His Val Thr Met Trp Val Asp Asp Gln Glu Val Ile Asn
115 120 125

aaa gct tct aat tct aac aaa atc aga tta gaa aaa gga aga tta tat 2232
Lys Ala Ser Asn Ser Asn Lys Ile Arg Leu Glu Lys Gly Arg Leu Tyr
130 135 140

caa ata aaa att caa tat caa cga gaa aat cct act gaa aaa gga ttg 2280
Gln Ile Lys Ile Gln Tyr Gln Arg Glu Asn Pro Thr Glu Lys Gly Leu
145 150 155

gat ttc aag ttg tac tgg acc gat tct caa aat aaa aaa gaa gtg att 2328
Asp Phe Lys Leu Tyr Trp Thr Asp Ser Gln Asn Lys Lys Glu Val Ile
160 165 170 175

tct agt gat aac tta caa ttg cca gaa tta aaa caa aaa tct tcg aac 2376
Ser Ser Asp Asn Leu Gln Leu Pro Glu Leu Lys Gln Lys Ser Ser Asn

```
                  180                     185                     190

tca aga aaa aag cga agt aca agt gct gga cct acg gtt cca gac cgt       2424
Ser Arg Lys Lys Arg Ser Thr Ser Ala Gly Pro Thr Val Pro Asp Arg
            195                     200                     205

gac aat gat gga atc cct gat tca tta gag gta gaa gga tat acg gtt       2472
Asp Asn Asp Gly Ile Pro Asp Ser Leu Glu Val Glu Gly Tyr Thr Val
        210                     215                     220

gat gtc aaa aat aaa aga act ttt ctt tca cca tgg att tct aat att       2520
Asp Val Lys Asn Lys Arg Thr Phe Leu Ser Pro Trp Ile Ser Asn Ile
    225                     230                     235

cat gaa aag aaa gga tta acc aaa tat aaa tca tct cct gaa aaa tgg       2568
His Glu Lys Lys Gly Leu Thr Lys Tyr Lys Ser Ser Pro Glu Lys Trp
240                 245                     250                 255

agc acg gct tct gat ccg tac agt gat ttc gaa aag gtt aca gga cgg       2616
Ser Thr Ala Ser Asp Pro Tyr Ser Asp Phe Glu Lys Val Thr Gly Arg
                260                     265                     270

att gat aag aat gta tca cca gag gca aga cac ccc ctt gtg gca gct       2664
Ile Asp Lys Asn Val Ser Pro Glu Ala Arg His Pro Leu Val Ala Ala
            275                     280                     285

tat ccg att gta cat gta gat atg gag aat att att ctc tca aaa aat       2712
Tyr Pro Ile Val His Val Asp Met Glu Asn Ile Ile Leu Ser Lys Asn
        290                     295                     300

gag gat caa tcc aca cag aat act gat agt gaa acg aga aca ata agt       2760
Glu Asp Gln Ser Thr Gln Asn Thr Asp Ser Glu Thr Arg Thr Ile Ser
    305                     310                     315

aaa aat act tct aca agt agg aca cat act agt gaa gta cat gga aat       2808
Lys Asn Thr Ser Thr Ser Arg Thr His Thr Ser Glu Val His Gly Asn
320                 325                     330                 335

gca gaa gtg cat gcg tcg ttc ttt gat att ggt ggg agt gta tct gca       2856
Ala Glu Val His Ala Ser Phe Phe Asp Ile Gly Gly Ser Val Ser Ala
                340                     345                     350

gga ttt agt aat tcg aat tca agt acg gtc gca att gat cat tca cta       2904
Gly Phe Ser Asn Ser Asn Ser Ser Thr Val Ala Ile Asp His Ser Leu
            355                     360                     365

tct cta gca ggg gaa aga act tgg gct gaa aca atg ggt tta aat acc       2952
Ser Leu Ala Gly Glu Arg Thr Trp Ala Glu Thr Met Gly Leu Asn Thr
        370                     375                     380

gct gat aca gca aga tta aat gcc aat att aga tat gta aat act ggg       3000
Ala Asp Thr Ala Arg Leu Asn Ala Asn Ile Arg Tyr Val Asn Thr Gly
    385                     390                     395

acg gct cca atc tac aac gtg tta cca acg act tcg tta gtg tta gga       3048
Thr Ala Pro Ile Tyr Asn Val Leu Pro Thr Thr Ser Leu Val Leu Gly
400                 405                     410                 415

aaa aat caa aca ctc gcg aca att aaa gct aag gaa aac caa tta agt       3096
Lys Asn Gln Thr Leu Ala Thr Ile Lys Ala Lys Glu Asn Gln Leu Ser
                420                     425                     430
```

caa ata ctt gca cct aat aat tat tat cct tct aaa aac ttg gcg cca 3144
Gln Ile Leu Ala Pro Asn Asn Tyr Tyr Pro Ser Lys Asn Leu Ala Pro
435 440 445

atc gca tta aat gca caa gac gat ttc agt tct act cca att aca atg 3192
Ile Ala Leu Asn Ala Gln Asp Asp Phe Ser Ser Thr Pro Ile Thr Met
450 455 460

aat tac aat caa ttt ctt gag tta gaa aaa acg aaa caa tta aga tta 3240
Asn Tyr Asn Gln Phe Leu Glu Leu Glu Lys Thr Lys Gln Leu Arg Leu
465 470 475

gat acg gat caa gta tat ggg aat ata gca aca tac aat ttt gaa aat 3288
Asp Thr Asp Gln Val Tyr Gly Asn Ile Ala Thr Tyr Asn Phe Glu Asn
480 485 490 495

gga aga gtg agg gtg gat aca ggc tcg aac tgg agt gaa gtg tta ccg 3336
Gly Arg Val Arg Val Asp Thr Gly Ser Asn Trp Ser Glu Val Leu Pro
500 505 510

caa att caa gaa aca act gca cgt atc att ttt aat gga aaa gat tta 3384
Gln Ile Gln Glu Thr Thr Ala Arg Ile Ile Phe Asn Gly Lys Asp Leu
515 520 525

aat ctg gta gaa agg cgg ata gcg gcg gtt aat cct agt gat cca tta 3432
Asn Leu Val Glu Arg Arg Ile Ala Ala Val Asn Pro Ser Asp Pro Leu
530 535 540

gaa acg act aaa ccg gat atg aca tta aaa gaa gcc ctt aaa ata gca 3480
Glu Thr Thr Lys Pro Asp Met Thr Leu Lys Glu Ala Leu Lys Ile Ala
545 550 555

ttt gga ttt aac gaa ccg aat gga aac tta caa tat caa ggg aaa gac 3528
Phe Gly Phe Asn Glu Pro Asn Gly Asn Leu Gln Tyr Gln Gly Lys Asp
560 565 570 575

ata acc gaa ttt gat ttt aat ttc gat caa caa aca tct caa aat atc 3576
Ile Thr Glu Phe Asp Phe Asn Phe Asp Gln Gln Thr Ser Gln Asn Ile
580 585 590

aag aat cag tta gcg gaa tta aac gca act aac ata tat act gta tta 3624
Lys Asn Gln Leu Ala Glu Leu Asn Ala Thr Asn Ile Tyr Thr Val Leu
595 600 605

gat aaa atc aaa tta aat gca aaa atg aat att tta ata aga gat aaa 3672
Asp Lys Ile Lys Leu Asn Ala Lys Met Asn Ile Leu Ile Arg Asp Lys
610 615 620

cgt ttt cat tat gat aga aat aac ata gca gtt ggg gcg gat gag tca 3720
Arg Phe His Tyr Asp Arg Asn Asn Ile Ala Val Gly Ala Asp Glu Ser
625 630 635

gta gtt aag gag gct cat aga gaa gta att aat tcg tca aca gag gga 3768
Val Val Lys Glu Ala His Arg Glu Val Ile Asn Ser Ser Thr Glu Gly
640 645 650 655

tta ttg tta aat att gat aag gat ata aga aaa ata tta tca ggt tat 3816
Leu Leu Leu Asn Ile Asp Lys Asp Ile Arg Lys Ile Leu Ser Gly Tyr
660 665 670

att gta gaa att gaa gat act gaa ggg ctt aaa gaa gtt ata aat gac 3864
Ile Val Glu Ile Glu Asp Thr Glu Gly Leu Lys Glu Val Ile Asn Asp

```
              675                 680                 685

aga tat gat atg ttg aat att tct agt tta cgg caa gat gga aaa aca      3912
Arg Tyr Asp Met Leu Asn Ile Ser Ser Leu Arg Gln Asp Gly Lys Thr
        690                 695                 700

ttt ata gat ttt aaa aaa tat aat gat aaa tta ccg tta tat ata agt      3960
Phe Ile Asp Phe Lys Lys Tyr Asn Asp Lys Leu Pro Leu Tyr Ile Ser
    705                 710                 715

aat ccc aat tat aag gta aat gta tat gct gtt act aaa gaa aac act      4008
Asn Pro Asn Tyr Lys Val Asn Val Tyr Ala Val Thr Lys Glu Asn Thr
720                 725                 730                 735

att att aat cct agt gag aat ggg gat act agt acc aac ggg atc aag      4056
Ile Ile Asn Pro Ser Glu Asn Gly Asp Thr Ser Thr Asn Gly Ile Lys
                740                 745                 750

aaa att tta atc ttt tct aaa aaa ggc tat gag ata gga taa              4098
Lys Ile Leu Ile Phe Ser Lys Lys Gly Tyr Glu Ile Gly
            755                 760

ggtaattcta ggtgattttt aaattatcta aaaaacagta aaattaaaac atactctttt   4158

tgtaagaaat acaaggagag tatgttttaa acagtaatct aaatcatcat aatcctttga    4218

gattgtttgt aggatcc                                                   4235


<210>  4

<211>  764

<212>  PRT

<213>  Bacillus anthracis


<400>  4

Met Lys Lys Arg Lys Val Leu Ile Pro Leu Met Ala Leu Ser Thr Ile
1               5                   10                  15

Leu Val Ser Ser Thr Gly Asn Leu Glu Val Ile Gln Ala Glu Val Lys
            20                  25                  30

Gln Glu Asn Arg Leu Leu Asn Glu Ser Glu Ser Ser Ser Gln Gly Leu
        35                  40                  45

Leu Gly Tyr Tyr Phe Ser Asp Leu Asn Phe Gln Ala Pro Met Val Val
    50                  55                  60

Thr Ser Ser Thr Thr Gly Asp Leu Ser Ile Pro Ser Ser Glu Leu Glu
65                  70                  75                  80

Asn Ile Pro Ser Glu Asn Gln Tyr Phe Gln Ser Ala Ile Trp Ser Gly
```

85 90 95

Phe Ile Lys Val Lys Lys Ser Asp Glu Tyr Thr Phe Ala Thr Ser Ala
100 105 110

Asp Asn His Val Thr Met Trp Val Asp Asp Gln Glu Val Ile Asn Lys
115 120 125

Ala Ser Asn Ser Asn Lys Ile Arg Leu Glu Lys Gly Arg Leu Tyr Gln
130 135 140

Ile Lys Ile Gln Tyr Gln Arg Glu Asn Pro Thr Glu Lys Gly Leu Asp
145 150 155 160

Phe Lys Leu Tyr Trp Thr Asp Ser Gln Asn Lys Lys Glu Val Ile Ser
165 170 175

Ser Asp Asn Leu Gln Leu Pro Glu Leu Lys Gln Lys Ser Ser Asn Ser
180 185 190

Arg Lys Lys Arg Ser Thr Ser Ala Gly Pro Thr Val Pro Asp Arg Asp
195 200 205

Asn Asp Gly Ile Pro Asp Ser Leu Glu Val Glu Gly Tyr Thr Val Asp
210 215 220

Val Lys Asn Lys Arg Thr Phe Leu Ser Pro Trp Ile Ser Asn Ile His
225 230 235 240

Glu Lys Lys Gly Leu Thr Lys Tyr Lys Ser Ser Pro Glu Lys Trp Ser
245 250 255

Thr Ala Ser Asp Pro Tyr Ser Asp Phe Glu Lys Val Thr Gly Arg Ile
260 265 270

Asp Lys Asn Val Ser Pro Glu Ala Arg His Pro Leu Val Ala Ala Tyr
275 280 285

Pro Ile Val His Val Asp Met Glu Asn Ile Ile Leu Ser Lys Asn Glu
290 295 300

Asp Gln Ser Thr Gln Asn Thr Asp Ser Glu Thr Arg Thr Ile Ser Lys
305 310 315 320

Asn Thr Ser Thr Ser Arg Thr His Thr Ser Glu Val His Gly Asn Ala
325 330 335

Glu Val His Ala Ser Phe Phe Asp Ile Gly Gly Ser Val Ser Ala Gly
340 345 350

Phe Ser Asn Ser Asn Ser Ser Thr Val Ala Ile Asp His Ser Leu Ser
355 360 365

Leu Ala Gly Glu Arg Thr Trp Ala Glu Thr Met Gly Leu Asn Thr Ala
370 375 380

Asp Thr Ala Arg Leu Asn Ala Asn Ile Arg Tyr Val Asn Thr Gly Thr
385 390 395 400

Ala Pro Ile Tyr Asn Val Leu Pro Thr Thr Ser Leu Val Leu Gly Lys
405 410 415

Asn Gln Thr Leu Ala Thr Ile Lys Ala Lys Glu Asn Gln Leu Ser Gln
420 425 430

Ile Leu Ala Pro Asn Asn Tyr Tyr Pro Ser Lys Asn Leu Ala Pro Ile
435 440 445

Ala Leu Asn Ala Gln Asp Asp Phe Ser Ser Thr Pro Ile Thr Met Asn
450 455 460

Tyr Asn Gln Phe Leu Glu Leu Glu Lys Thr Lys Gln Leu Arg Leu Asp
465 470 475 480

Thr Asp Gln Val Tyr Gly Asn Ile Ala Thr Tyr Asn Phe Glu Asn Gly
485 490 495

Arg Val Arg Val Asp Thr Gly Ser Asn Trp Ser Glu Val Leu Pro Gln
500 505 510

Ile Gln Glu Thr Thr Ala Arg Ile Ile Phe Asn Gly Lys Asp Leu Asn
515 520 525

Leu Val Glu Arg Arg Ile Ala Ala Val Asn Pro Ser Asp Pro Leu Glu
530 535 540

Thr Thr Lys Pro Asp Met Thr Leu Lys Glu Ala Leu Lys Ile Ala Phe
545 550 555 560

Gly Phe Asn Glu Pro Asn Gly Asn Leu Gln Tyr Gln Gly Lys Asp Ile
565 570 575

Thr Glu Phe Asp Phe Asn Phe Asp Gln Gln Thr Ser Gln Asn Ile Lys
580 585 590

```
Asn Gln Leu Ala Glu Leu Asn Ala Thr Asn Ile Tyr Thr Val Leu Asp
        595                 600                 605

Lys Ile Lys Leu Asn Ala Lys Met Asn Ile Leu Ile Arg Asp Lys Arg
    610                 615                 620

Phe His Tyr Asp Arg Asn Asn Ile Ala Val Gly Ala Asp Glu Ser Val
625                 630                 635                 640

Val Lys Glu Ala His Arg Glu Val Ile Asn Ser Ser Thr Glu Gly Leu
                645                 650                 655

Leu Leu Asn Ile Asp Lys Asp Ile Arg Lys Ile Leu Ser Gly Tyr Ile
            660                 665                 670

Val Glu Ile Glu Asp Thr Glu Gly Leu Lys Glu Val Ile Asn Asp Arg
        675                 680                 685

Tyr Asp Met Leu Asn Ile Ser Ser Leu Arg Gln Asp Gly Lys Thr Phe
    690                 695                 700

Ile Asp Phe Lys Lys Tyr Asn Asp Lys Leu Pro Leu Tyr Ile Ser Asn
705                 710                 715                 720

Pro Asn Tyr Lys Val Asn Val Tyr Ala Val Thr Lys Glu Asn Thr Ile
                725                 730                 735

Ile Asn Pro Ser Glu Asn Gly Asp Thr Ser Thr Asn Gly Ile Lys Lys
            740                 745                 750

Ile Leu Ile Phe Ser Lys Lys Gly Tyr Glu Ile Gly
        755                 760
```

<210> 5

<211> 1782

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic coding region for Human TPA/synthetic antigen fusion protein

<220>

```
<221>  CDS

<222>  (13)..(1773)

<223>



<400>  5
gatatcgcca cc atg gat gca atg aag aga ggg ctc tgc tgt gtg ctg ctg       51
              Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu
              1               5                   10

ctg tgt gga gca gtc ttc gtt tcg ccc agc agc gct ggg cca act gtg         99
Leu Cys Gly Ala Val Phe Val Ser Pro Ser Ser Ala Gly Pro Thr Val
    15                  20                  25

ccc gac aga gac aat gat gga atc cct gat agt cta gag gtt gag gga        147
Pro Asp Arg Asp Asn Asp Gly Ile Pro Asp Ser Leu Glu Val Glu Gly
30                  35                  40                  45

tac acg gta gat gtc aag aac aaa agg act ttt ctc tcg cct tgg atc        195
Tyr Thr Val Asp Val Lys Asn Lys Arg Thr Phe Leu Ser Pro Trp Ile
                50                  55                  60

tca aat atc cat gag aag aag ggg ctt acc aag tac aag tcc tcc ccc        243
Ser Asn Ile His Glu Lys Lys Gly Leu Thr Lys Tyr Lys Ser Ser Pro
            65                  70                  75

gag aag tgg tct acc gct tcc gat cca tat agc gat ttc gag aag gtc        291
Glu Lys Trp Ser Thr Ala Ser Asp Pro Tyr Ser Asp Phe Glu Lys Val
        80                  85                  90

aca ggc cgg atc gat aaa aat gtg tct cca gag gct aga cac ccc ctg        339
Thr Gly Arg Ile Asp Lys Asn Val Ser Pro Glu Ala Arg His Pro Leu
    95                  100                 105

gta gca gcc tac ccg att gta cac gtg gac atg gag aac atc att cta        387
Val Ala Ala Tyr Pro Ile Val His Val Asp Met Glu Asn Ile Ile Leu
110                 115                 120                 125

agc aaa aac gag gac cag tcc aca caa aac act gac tcc gag acc cgc        435
Ser Lys Asn Glu Asp Gln Ser Thr Gln Asn Thr Asp Ser Glu Thr Arg
                130                 135                 140

acc ata tct aaa aac acc agt act tca agg acc cac acc tct gaa gtg        483
Thr Ile Ser Lys Asn Thr Ser Thr Ser Arg Thr His Thr Ser Glu Val
            145                 150                 155

cac ggc aat gcg gaa gtc cat gca tcg gat att ggt ggc tcc gtg tca        531
His Gly Asn Ala Glu Val His Ala Ser Asp Ile Gly Gly Ser Val Ser
        160                 165                 170

gcc ggc ttt agc aat agc aac tcc tcg acg gtt gcc att gac cac tca        579
Ala Gly Phe Ser Asn Ser Asn Ser Ser Thr Val Ala Ile Asp His Ser
    175                 180                 185

ctg tca tta gca ggt gag agg act tgg gct gaa act atg ggt ctg aat        627
Leu Ser Leu Ala Gly Glu Arg Thr Trp Ala Glu Thr Met Gly Leu Asn
190                 195                 200                 205

acc gcc gat acg gcc cgg ctc aac gca aat att cgg tac gtc aac aca        675
```

```
Thr Ala Asp Thr Ala Arg Leu Asn Ala Asn Ile Arg Tyr Val Asn Thr
                210                 215                 220

ggg act gct cct ata tat aac gtg ctg cct acg aca agt ctt gtc ctg      723
Gly Thr Ala Pro Ile Tyr Asn Val Leu Pro Thr Thr Ser Leu Val Leu
            225                 230                 235

ggc aaa aat cag acc ctc gca acc att aag gca aag gaa aat cag ctg      771
Gly Lys Asn Gln Thr Leu Ala Thr Ile Lys Ala Lys Glu Asn Gln Leu
        240                 245                 250

agc cag atc ctc gcc cct aac aac tat tat cca tcc aaa aat tta gcc      819
Ser Gln Ile Leu Ala Pro Asn Asn Tyr Tyr Pro Ser Lys Asn Leu Ala
    255                 260                 265

ccc ata gcc ctg aac gcc cag gac gac ttt tcc tct acc ccc ata act      867
Pro Ile Ala Leu Asn Ala Gln Asp Asp Phe Ser Ser Thr Pro Ile Thr
270                 275                 280                 285

atg aat tac aat cag ttc ctg gag ctg gaa aag acg aag cag ctg aga      915
Met Asn Tyr Asn Gln Phe Leu Glu Leu Glu Lys Thr Lys Gln Leu Arg
                290                 295                 300

cta gac acc gat cag gtg tat gga aac ata gcg aca tat aac ttt gag      963
Leu Asp Thr Asp Gln Val Tyr Gly Asn Ile Ala Thr Tyr Asn Phe Glu
            305                 310                 315

aac ggc cgc gtg cgc gtc gac act ggg tca aac tgg tct gaa gtt ctg     1011
Asn Gly Arg Val Arg Val Asp Thr Gly Ser Asn Trp Ser Glu Val Leu
        320                 325                 330

ccg caa att caa gag aca acc gcc aga att atc ttt aat ggg aag gac     1059
Pro Gln Ile Gln Glu Thr Thr Ala Arg Ile Ile Phe Asn Gly Lys Asp
    335                 340                 345

ttg aac ctt gtc gaa cgt aga att gcc gcc gtg aac ccc agt gat cca     1107
Leu Asn Leu Val Glu Arg Arg Ile Ala Ala Val Asn Pro Ser Asp Pro
350                 355                 360                 365

ctc gag acg act aaa ccg gat atg aca ctg aaa gag gct ctg aag att     1155
Leu Glu Thr Thr Lys Pro Asp Met Thr Leu Lys Glu Ala Leu Lys Ile
                370                 375                 380

gcc ttc gga ttc aac gaa cct aat ggc aat ttg cag tat cag ggg aaa     1203
Ala Phe Gly Phe Asn Glu Pro Asn Gly Asn Leu Gln Tyr Gln Gly Lys
            385                 390                 395

gac atc aca gag ttt gat ttc aat ttc gat cag cag act tcc caa aat     1251
Asp Ile Thr Glu Phe Asp Phe Asn Phe Asp Gln Gln Thr Ser Gln Asn
        400                 405                 410

atc aaa aat cag ttg gca gag ctg aat gcc acc aat atc tac acg gtt     1299
Ile Lys Asn Gln Leu Ala Glu Leu Asn Ala Thr Asn Ile Tyr Thr Val
    415                 420                 425

ctc gat aaa atc aaa ctt aac gcc aag atg aac ata ttg att cga gac     1347
Leu Asp Lys Ile Lys Leu Asn Ala Lys Met Asn Ile Leu Ile Arg Asp
430                 435                 440                 445

aaa cgc ttc cac tac gac cgc aac aat ata gcc gta ggc gct gat gag     1395
Lys Arg Phe His Tyr Asp Arg Asn Asn Ile Ala Val Gly Ala Asp Glu
                450                 455                 460
```

```
tct gtc gtc aag gag gct cat agg gaa gtt atc aac agc agt act gaa      1443
Ser Val Val Lys Glu Ala His Arg Glu Val Ile Asn Ser Ser Thr Glu
            465                 470                 475

ggg ctg tta ctt aat atc gac aag gac att cgg aag atc ctg tcc ggg      1491
Gly Leu Leu Leu Asn Ile Asp Lys Asp Ile Arg Lys Ile Leu Ser Gly
        480                 485                 490

tat atc gtg gag atc gag gat acc gag ggc ctg aag gaa gtc att aac      1539
Tyr Ile Val Glu Ile Glu Asp Thr Glu Gly Leu Lys Glu Val Ile Asn
    495                 500                 505

gac cgc tat gat atg ctg aac att tcc agc tta cga cag gac ggt aag      1587
Asp Arg Tyr Asp Met Leu Asn Ile Ser Ser Leu Arg Gln Asp Gly Lys
510                 515                 520                 525

aca ttt att gac ttt aaa aag tat aac gac aag cta ccc ctg tac att      1635
Thr Phe Ile Asp Phe Lys Lys Tyr Asn Asp Lys Leu Pro Leu Tyr Ile
                530                 535                 540

tcc aac cca aat tac aaa gtt aat gtg tat gct gta acc aag gag aac      1683
Ser Asn Pro Asn Tyr Lys Val Asn Val Tyr Ala Val Thr Lys Glu Asn
            545                 550                 555

aca atc atc aat cca agc gag aac ggc gat acc agc aca aat gga atc      1731
Thr Ile Ile Asn Pro Ser Glu Asn Gly Asp Thr Ser Thr Asn Gly Ile
        560                 565                 570

aaa aag atc ctt ata ttt agt aaa aaa ggc tac gag atc ggt tgaggatcc    1782
Lys Lys Ile Leu Ile Phe Ser Lys Lys Gly Tyr Glu Ile Gly
    575                 580                 585
```

<210> 6

<211> 587

<212> PRT

<213> Artificial Sequence

<220>

<223> Human TPA/synthetic antigen fusion protein

<400> 6

```
Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
1               5                   10                  15

Ala Val Phe Val Ser Pro Ser Ser Ala Gly Pro Thr Val Pro Asp Arg
            20                  25                  30

Asp Asn Asp Gly Ile Pro Asp Ser Leu Glu Val Glu Gly Tyr Thr Val
        35                  40                  45

Asp Val Lys Asn Lys Arg Thr Phe Leu Ser Pro Trp Ile Ser Asn Ile
```

50 55 60

His Glu Lys Lys Gly Leu Thr Lys Tyr Lys Ser Ser Pro Glu Lys Trp
65 70 75 80

Ser Thr Ala Ser Asp Pro Tyr Ser Asp Phe Glu Lys Val Thr Gly Arg
85 90 95

Ile Asp Lys Asn Val Ser Pro Glu Ala Arg His Pro Leu Val Ala Ala
100 105 110

Tyr Pro Ile Val His Val Asp Met Glu Asn Ile Ile Leu Ser Lys Asn
115 120 125

Glu Asp Gln Ser Thr Gln Asn Thr Asp Ser Glu Thr Arg Thr Ile Ser
130 135 140

Lys Asn Thr Ser Thr Ser Arg Thr His Thr Ser Glu Val His Gly Asn
145 150 155 160

Ala Glu Val His Ala Ser Asp Ile Gly Gly Ser Val Ser Ala Gly Phe
165 170 175

Ser Asn Ser Asn Ser Ser Thr Val Ala Ile Asp His Ser Leu Ser Leu
180 185 190

Ala Gly Glu Arg Thr Trp Ala Glu Thr Met Gly Leu Asn Thr Ala Asp
195 200 205

Thr Ala Arg Leu Asn Ala Asn Ile Arg Tyr Val Asn Thr Gly Thr Ala
210 215 220

Pro Ile Tyr Asn Val Leu Pro Thr Thr Ser Leu Val Leu Gly Lys Asn
225 230 235 240

Gln Thr Leu Ala Thr Ile Lys Ala Lys Glu Asn Gln Leu Ser Gln Ile
245 250 255

Leu Ala Pro Asn Asn Tyr Tyr Pro Ser Lys Asn Leu Ala Pro Ile Ala
260 265 270

Leu Asn Ala Gln Asp Asp Phe Ser Ser Thr Pro Ile Thr Met Asn Tyr
275 280 285

Asn Gln Phe Leu Glu Leu Glu Lys Thr Lys Gln Leu Arg Leu Asp Thr
290 295 300

```
Asp Gln Val Tyr Gly Asn Ile Ala Thr Tyr Asn Phe Glu Asn Gly Arg
305                 310                 315                 320

Val Arg Val Asp Thr Gly Ser Asn Trp Ser Glu Val Leu Pro Gln Ile
                325                 330                 335

Gln Glu Thr Thr Ala Arg Ile Ile Phe Asn Gly Lys Asp Leu Asn Leu
            340                 345                 350

Val Glu Arg Arg Ile Ala Ala Val Asn Pro Ser Asp Pro Leu Glu Thr
        355                 360                 365

Thr Lys Pro Asp Met Thr Leu Lys Glu Ala Leu Lys Ile Ala Phe Gly
    370                 375                 380

Phe Asn Glu Pro Asn Gly Asn Leu Gln Tyr Gln Gly Lys Asp Ile Thr
385                 390                 395                 400

Glu Phe Asp Phe Asn Phe Asp Gln Gln Thr Ser Gln Asn Ile Lys Asn
                405                 410                 415

Gln Leu Ala Glu Leu Asn Ala Thr Asn Ile Tyr Thr Val Leu Asp Lys
            420                 425                 430

Ile Lys Leu Asn Ala Lys Met Asn Ile Leu Ile Arg Asp Lys Arg Phe
        435                 440                 445

His Tyr Asp Arg Asn Asn Ile Ala Val Gly Ala Asp Glu Ser Val Val
    450                 455                 460

Lys Glu Ala His Arg Glu Val Ile Asn Ser Ser Thr Glu Gly Leu Leu
465                 470                 475                 480

Leu Asn Ile Asp Lys Asp Ile Arg Lys Ile Leu Ser Gly Tyr Ile Val
                485                 490                 495

Glu Ile Glu Asp Thr Glu Gly Leu Lys Glu Val Ile Asn Asp Arg Tyr
            500                 505                 510

Asp Met Leu Asn Ile Ser Ser Leu Arg Gln Asp Gly Lys Thr Phe Ile
        515                 520                 525

Asp Phe Lys Lys Tyr Asn Asp Lys Leu Pro Leu Tyr Ile Ser Asn Pro
    530                 535                 540

Asn Tyr Lys Val Asn Val Tyr Ala Val Thr Lys Glu Asn Thr Ile Ile
545                 550                 555                 560
```

```
Asn Pro Ser Glu Asn Gly Asp Thr Ser Thr Asn Gly Ile Lys Lys Ile
                565                 570                 575

Leu Ile Phe Ser Lys Lys Gly Tyr Glu Ile Gly
            580                 585
```

<210> 7

<211> 2277

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic coding region for Human TPA/synthetic antigen fusion protein

<220>

<221> CDS

<222> (13)..(2268)

<223>

<400> 7

```
gatatcgcca cc atg gat gca atg aag aga ggg ctc tgc tgt gtg ctg ctg      51
              Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu
              1               5                   10

ctg tgt gga gca gtc ttc gtt tcg ccc agc gaa gtg aag caa gaa aat        99
Leu Cys Gly Ala Val Phe Val Ser Pro Ser Glu Val Lys Gln Glu Asn
    15                  20                  25

cga ctt ctg aac gag agc gaa agt tca tca cag ggt ctt ctc gga tac       147
Arg Leu Leu Asn Glu Ser Glu Ser Ser Ser Gln Gly Leu Leu Gly Tyr
30                  35                  40                  45

tac ttc agt gac ttg aat ttc caa gca cca atg gtg gtg act agt agc       195
Tyr Phe Ser Asp Leu Asn Phe Gln Ala Pro Met Val Val Thr Ser Ser
                50                  55                  60

acc acc ggc gat ttg agc att ccc agc tct gag ttg gag aac att ccc       243
Thr Thr Gly Asp Leu Ser Ile Pro Ser Ser Glu Leu Glu Asn Ile Pro
            65                  70                  75

agc gaa aat cag tac ttc cag tct gct atc tgg tcc gga ttc att aag       291
Ser Glu Asn Gln Tyr Phe Gln Ser Ala Ile Trp Ser Gly Phe Ile Lys
        80                  85                  90

gtt aaa aag tcc gac gaa tat aca ttt gct acc tcg gcg gat aac cat       339
Val Lys Lys Ser Asp Glu Tyr Thr Phe Ala Thr Ser Ala Asp Asn His
    95                  100                 105
```

```
gtg aca atg tgg gtg gac gac cag gaa gtg atc aac aag gct tca aac        387
Val Thr Met Trp Val Asp Asp Gln Glu Val Ile Asn Lys Ala Ser Asn
110                 115                 120                 125

tct aat aaa atc cgg ctc gag aag ggg agg ctc tac cag atc aaa att        435
Ser Asn Lys Ile Arg Leu Glu Lys Gly Arg Leu Tyr Gln Ile Lys Ile
                130                 135                 140

cag tac cag cgg gaa aac cct aca gaa aaa gga ctc gat ttc aag ctg        483
Gln Tyr Gln Arg Glu Asn Pro Thr Glu Lys Gly Leu Asp Phe Lys Leu
            145                 150                 155

tac tgg aca gat agc caa aac aag aaa gaa gtt atc agc tca gac aat        531
Tyr Trp Thr Asp Ser Gln Asn Lys Lys Glu Val Ile Ser Ser Asp Asn
        160                 165                 170

ctg cag tta ccc gag ctc aag cag aag agt tct aat aca agc gct ggg        579
Leu Gln Leu Pro Glu Leu Lys Gln Lys Ser Ser Asn Thr Ser Ala Gly
    175                 180                 185

cca act gtg ccc gac aga gac aat gat gga atc cct gat agt cta gag        627
Pro Thr Val Pro Asp Arg Asp Asn Asp Gly Ile Pro Asp Ser Leu Glu
190                 195                 200                 205

gtt gag gga tac acg gta gat gtc aag aac aaa agg act ttt ctc tcg        675
Val Glu Gly Tyr Thr Val Asp Val Lys Asn Lys Arg Thr Phe Leu Ser
                210                 215                 220

cct tgg atc tca aat atc cat gag aag aag ggg ctt acc aag tac aag        723
Pro Trp Ile Ser Asn Ile His Glu Lys Lys Gly Leu Thr Lys Tyr Lys
            225                 230                 235

tcc tcc ccc gag aag tgg tct acc gct tcc gat cca tat agc gat ttc        771
Ser Ser Pro Glu Lys Trp Ser Thr Ala Ser Asp Pro Tyr Ser Asp Phe
        240                 245                 250

gag aag gtc aca ggc cgg atc gat aaa aat gtg tct cca gag gct aga        819
Glu Lys Val Thr Gly Arg Ile Asp Lys Asn Val Ser Pro Glu Ala Arg
    255                 260                 265

cac ccc ctg gta gca gcc tac ccg att gta cac gtg gac atg gag aac        867
His Pro Leu Val Ala Ala Tyr Pro Ile Val His Val Asp Met Glu Asn
270                 275                 280                 285

atc att cta agc aaa aac gag gac cag tcc aca caa aac act gac tcc        915
Ile Ile Leu Ser Lys Asn Glu Asp Gln Ser Thr Gln Asn Thr Asp Ser
                290                 295                 300

gag acc cgc acc ata tct aaa aac acc agt act tca agg acc cac acc        963
Glu Thr Arg Thr Ile Ser Lys Asn Thr Ser Thr Ser Arg Thr His Thr
            305                 310                 315

tct gaa gtg cac ggc aat gcg gaa gtc cat gca tcg ttt ttc gat att       1011
Ser Glu Val His Gly Asn Ala Glu Val His Ala Ser Phe Phe Asp Ile
        320                 325                 330

ggt ggc tcc gtg tca gcc ggc ttt agc aat agc aac tcc tcg acg gtt       1059
Gly Gly Ser Val Ser Ala Gly Phe Ser Asn Ser Asn Ser Ser Thr Val
    335                 340                 345

gcc att gac cac tca ctg tca tta gca ggt gag agg act tgg gct gaa       1107
Ala Ile Asp His Ser Leu Ser Leu Ala Gly Glu Arg Thr Trp Ala Glu
```

```
350                 355                 360                 365

act atg ggt ctg aat acc gcc gat acg gcc cgg ctc aac gca aat att      1155
Thr Met Gly Leu Asn Thr Ala Asp Thr Ala Arg Leu Asn Ala Asn Ile
                370                 375                 380

cgg tac gtc aac aca ggg act gct cct ata tat aac gtg ctg cct acg      1203
Arg Tyr Val Asn Thr Gly Thr Ala Pro Ile Tyr Asn Val Leu Pro Thr
            385                 390                 395

aca agt ctt gtc ctg ggc aaa aat cag acc ctc gca acc att aag gca      1251
Thr Ser Leu Val Leu Gly Lys Asn Gln Thr Leu Ala Thr Ile Lys Ala
        400                 405                 410

aag gaa aat cag ctg agc cag atc ctc gcc cct aac aac tat tat cca      1299
Lys Glu Asn Gln Leu Ser Gln Ile Leu Ala Pro Asn Asn Tyr Tyr Pro
    415                 420                 425

tcc aaa aat tta gcc ccc ata gcc ctg aac gcc cag gac gac ttt tcc      1347
Ser Lys Asn Leu Ala Pro Ile Ala Leu Asn Ala Gln Asp Asp Phe Ser
430                 435                 440                 445

tct acc ccc ata act atg aat tac aat cag ttc ctg gag ctg gaa aag      1395
Ser Thr Pro Ile Thr Met Asn Tyr Asn Gln Phe Leu Glu Leu Glu Lys
                450                 455                 460

acg aag cag ctg aga cta gac acc gat cag gtg tat gga aac ata gcg      1443
Thr Lys Gln Leu Arg Leu Asp Thr Asp Gln Val Tyr Gly Asn Ile Ala
            465                 470                 475

aca tat aac ttt gag aac ggc cgc gtg cgc gtc gac act ggg tca aac      1491
Thr Tyr Asn Phe Glu Asn Gly Arg Val Arg Val Asp Thr Gly Ser Asn
        480                 485                 490

tgg tct gaa gtt ctg ccg caa att caa gag aca acc gcc aga att atc      1539
Trp Ser Glu Val Leu Pro Gln Ile Gln Glu Thr Thr Ala Arg Ile Ile
    495                 500                 505

ttt aat ggg aag gac ttg aac ctt gtc gaa cgt aga att gcc gcc gtg      1587
Phe Asn Gly Lys Asp Leu Asn Leu Val Glu Arg Arg Ile Ala Ala Val
510                 515                 520                 525

aac ccc agt gat cca ctc gag acg act aaa ccg gat atg aca ctg aaa      1635
Asn Pro Ser Asp Pro Leu Glu Thr Thr Lys Pro Asp Met Thr Leu Lys
                530                 535                 540

gag gct ctg aag att gcc ttc gga ttc aac gaa cct aat ggc aat ttg      1683
Glu Ala Leu Lys Ile Ala Phe Gly Phe Asn Glu Pro Asn Gly Asn Leu
            545                 550                 555

cag tat cag ggg aaa gac atc aca gag ttt gat ttc aat ttc gat cag      1731
Gln Tyr Gln Gly Lys Asp Ile Thr Glu Phe Asp Phe Asn Phe Asp Gln
        560                 565                 570

cag act tcc caa aat atc aaa aat cag ttg gca gag ctg aat gcc acc      1779
Gln Thr Ser Gln Asn Ile Lys Asn Gln Leu Ala Glu Leu Asn Ala Thr
    575                 580                 585

aat atc tac acg gtt ctc gat aaa atc aaa ctt aac gcc aag atg aac      1827
Asn Ile Tyr Thr Val Leu Asp Lys Ile Lys Leu Asn Ala Lys Met Asn
590                 595                 600                 605
```

```
ata ttg att cga gac aaa cgc ttc cac tac gac cgc aac aat ata gcc      1875
Ile Leu Ile Arg Asp Lys Arg Phe His Tyr Asp Arg Asn Asn Ile Ala
                610                 615                 620

gta ggc gct gat gag tct gtc gtc aag gag gct cat agg gaa gtt atc      1923
Val Gly Ala Asp Glu Ser Val Val Lys Glu Ala His Arg Glu Val Ile
            625                 630                 635

aac agc agt act gaa ggg ctg tta ctt aat atc gac aag gac att cgg      1971
Asn Ser Ser Thr Glu Gly Leu Leu Leu Asn Ile Asp Lys Asp Ile Arg
        640                 645                 650

aag atc ctg tcc ggg tat atc gtg gag atc gag gat acc gag ggc ctg      2019
Lys Ile Leu Ser Gly Tyr Ile Val Glu Ile Glu Asp Thr Glu Gly Leu
    655                 660                 665

aag gaa gtc att aac gac cgc tat gat atg ctg aac att tcc agc tta      2067
Lys Glu Val Ile Asn Asp Arg Tyr Asp Met Leu Asn Ile Ser Ser Leu
670                 675                 680                 685

cga cag gac ggt aag aca ttt att gac ttt aaa aag tat aac gac aag      2115
Arg Gln Asp Gly Lys Thr Phe Ile Asp Phe Lys Lys Tyr Asn Asp Lys
                690                 695                 700

cta ccc ctg tac att tcc aac cca aat tac aaa gtt aat gtg tat gct      2163
Leu Pro Leu Tyr Ile Ser Asn Pro Asn Tyr Lys Val Asn Val Tyr Ala
            705                 710                 715

gta acc aag gag aac aca atc atc aat cca agc gag aac ggc gat acc      2211
Val Thr Lys Glu Asn Thr Ile Ile Asn Pro Ser Glu Asn Gly Asp Thr
        720                 725                 730

agc aca aat gga atc aaa aag atc ctt ata ttt agt aaa aaa ggc tac      2259
Ser Thr Asn Gly Ile Lys Lys Ile Leu Ile Phe Ser Lys Lys Gly Tyr
    735                 740                 745

gag atc ggt tgaggatcc                                                2277
Glu Ile Gly
750
```

<210> 8

<211> 752

<212> PRT

<213> Artificial Sequence

<220>

<223> Human TPA/synthetic antigen fusion protein

<400> 8

```
Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
1               5                   10                  15

Ala Val Phe Val Ser Pro Ser Glu Val Lys Gln Glu Asn Arg Leu Leu
```

20 25 30

Asn Glu Ser Glu Ser Ser Ser Gln Gly Leu Leu Gly Tyr Tyr Phe Ser
35 40 45

Asp Leu Asn Phe Gln Ala Pro Met Val Val Thr Ser Ser Thr Thr Gly
50 55 60

Asp Leu Ser Ile Pro Ser Ser Glu Leu Glu Asn Ile Pro Ser Glu Asn
65 70 75 80

Gln Tyr Phe Gln Ser Ala Ile Trp Ser Gly Phe Ile Lys Val Lys Lys
85 90 95

Ser Asp Glu Tyr Thr Phe Ala Thr Ser Ala Asp Asn His Val Thr Met
100 105 110

Trp Val Asp Asp Gln Glu Val Ile Asn Lys Ala Ser Asn Ser Asn Lys
115 120 125

Ile Arg Leu Glu Lys Gly Arg Leu Tyr Gln Ile Lys Ile Gln Tyr Gln
130 135 140

Arg Glu Asn Pro Thr Glu Lys Gly Leu Asp Phe Lys Leu Tyr Trp Thr
145 150 155 160

Asp Ser Gln Asn Lys Lys Glu Val Ile Ser Ser Asp Asn Leu Gln Leu
165 170 175

Pro Glu Leu Lys Gln Lys Ser Ser Asn Thr Ser Ala Gly Pro Thr Val
180 185 190

Pro Asp Arg Asp Asn Asp Gly Ile Pro Asp Ser Leu Glu Val Glu Gly
195 200 205

Tyr Thr Val Asp Val Lys Asn Lys Arg Thr Phe Leu Ser Pro Trp Ile
210 215 220

Ser Asn Ile His Glu Lys Lys Gly Leu Thr Lys Tyr Lys Ser Ser Pro
225 230 235 240

Glu Lys Trp Ser Thr Ala Ser Asp Pro Tyr Ser Asp Phe Glu Lys Val
245 250 255

Thr Gly Arg Ile Asp Lys Asn Val Ser Pro Glu Ala Arg His Pro Leu
260 265 270

```
Val Ala Ala Tyr Pro Ile Val His Val Asp Met Glu Asn Ile Ile Leu
        275                 280                 285

Ser Lys Asn Glu Asp Gln Ser Thr Gln Asn Thr Asp Ser Glu Thr Arg
    290                 295                 300

Thr Ile Ser Lys Asn Thr Ser Thr Ser Arg Thr His Thr Ser Glu Val
305                 310                 315                 320

His Gly Asn Ala Glu Val His Ala Ser Phe Phe Asp Ile Gly Gly Ser
                325                 330                 335

Val Ser Ala Gly Phe Ser Asn Ser Asn Ser Ser Thr Val Ala Ile Asp
            340                 345                 350

His Ser Leu Ser Leu Ala Gly Glu Arg Thr Trp Ala Glu Thr Met Gly
        355                 360                 365

Leu Asn Thr Ala Asp Thr Ala Arg Leu Asn Ala Asn Ile Arg Tyr Val
    370                 375                 380

Asn Thr Gly Thr Ala Pro Ile Tyr Asn Val Leu Pro Thr Thr Ser Leu
385                 390                 395                 400

Val Leu Gly Lys Asn Gln Thr Leu Ala Thr Ile Lys Ala Lys Glu Asn
                405                 410                 415

Gln Leu Ser Gln Ile Leu Ala Pro Asn Asn Tyr Tyr Pro Ser Lys Asn
            420                 425                 430

Leu Ala Pro Ile Ala Leu Asn Ala Gln Asp Asp Phe Ser Ser Thr Pro
        435                 440                 445

Ile Thr Met Asn Tyr Asn Gln Phe Leu Glu Leu Glu Lys Thr Lys Gln
    450                 455                 460

Leu Arg Leu Asp Thr Asp Gln Val Tyr Gly Asn Ile Ala Thr Tyr Asn
465                 470                 475                 480

Phe Glu Asn Gly Arg Val Arg Val Asp Thr Gly Ser Asn Trp Ser Glu
                485                 490                 495

Val Leu Pro Gln Ile Gln Glu Thr Thr Ala Arg Ile Ile Phe Asn Gly
            500                 505                 510

Lys Asp Leu Asn Leu Val Glu Arg Arg Ile Ala Ala Val Asn Pro Ser
        515                 520                 525
```

```
Asp Pro Leu Glu Thr Thr Lys Pro Asp Met Thr Leu Lys Glu Ala Leu
    530                 535                 540

Lys Ile Ala Phe Gly Phe Asn Glu Pro Asn Gly Asn Leu Gln Tyr Gln
545                 550                 555                 560

Gly Lys Asp Ile Thr Glu Phe Asp Phe Asn Phe Asp Gln Gln Thr Ser
                565                 570                 575

Gln Asn Ile Lys Asn Gln Leu Ala Glu Leu Asn Ala Thr Asn Ile Tyr
            580                 585                 590

Thr Val Leu Asp Lys Ile Lys Leu Asn Ala Lys Met Asn Ile Leu Ile
        595                 600                 605

Arg Asp Lys Arg Phe His Tyr Asp Arg Asn Asn Ile Ala Val Gly Ala
    610                 615                 620

Asp Glu Ser Val Val Lys Glu Ala His Arg Glu Val Ile Asn Ser Ser
625                 630                 635                 640

Thr Glu Gly Leu Leu Leu Asn Ile Asp Lys Asp Ile Arg Lys Ile Leu
                645                 650                 655

Ser Gly Tyr Ile Val Glu Ile Glu Asp Thr Glu Gly Leu Lys Glu Val
            660                 665                 670

Ile Asn Asp Arg Tyr Asp Met Leu Asn Ile Ser Ser Leu Arg Gln Asp
        675                 680                 685

Gly Lys Thr Phe Ile Asp Phe Lys Lys Tyr Asn Asp Lys Leu Pro Leu
    690                 695                 700

Tyr Ile Ser Asn Pro Asn Tyr Lys Val Asn Val Tyr Ala Val Thr Lys
705                 710                 715                 720

Glu Asn Thr Ile Ile Asn Pro Ser Glu Asn Gly Asp Thr Ser Thr Asn
                725                 730                 735

Gly Ile Lys Lys Ile Leu Ile Phe Ser Lys Lys Gly Tyr Glu Ile Gly
            740                 745                 750
```

<210> 9

<211> 2418

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic coding region for Human TPA/synthetic
antigen fusion protein

<220>

<221> CDS

<222> (13)..(2409)

<223>

```
<400>  9
gatatcgcca cc atg gat gca atg aag aga ggg ctc tgc tgt gtg ctg ctg      51
              Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu
              1               5                   10

ctg tgt gga gca gtc ttc gtt tcg ccc agc gcc ggc ggg cat ggg gac        99
Leu Cys Gly Ala Val Phe Val Ser Pro Ser Ala Gly Gly His Gly Asp
    15                  20                  25

gtt ggc atg cat gtg aaa gaa aag gag aaa aac aag gac gaa aac aag       147
Val Gly Met His Val Lys Glu Lys Glu Lys Asn Lys Asp Glu Asn Lys
30                  35                  40                  45

cgt aaa gac gaa gaa cgt aat aaa aca cag gag gaa cac tta aag gag       195
Arg Lys Asp Glu Glu Arg Asn Lys Thr Gln Glu Glu His Leu Lys Glu
                50                  55                  60

atc atg aag cac ata gta aag att gag gta aaa ggc gaa gag gct gta       243
Ile Met Lys His Ile Val Lys Ile Glu Val Lys Gly Glu Glu Ala Val
            65                  70                  75

aag aag gag gca gca gaa aaa ctg ttg gag aag gtg cct tct gac gtc       291
Lys Lys Glu Ala Ala Glu Lys Leu Leu Glu Lys Val Pro Ser Asp Val
        80                  85                  90

tta gag atg tat aag gcc atc ggc ggt aag atc tat atc gtg gac gga       339
Leu Glu Met Tyr Lys Ala Ile Gly Gly Lys Ile Tyr Ile Val Asp Gly
    95                  100                 105

gac atc act aaa cac ata tct ctc gaa gct ctc tcc gag gac aag aaa       387
Asp Ile Thr Lys His Ile Ser Leu Glu Ala Leu Ser Glu Asp Lys Lys
110                 115                 120                 125

aag att aaa gac atc tac ggg aag gat gcc tta ttg cac gag cac tac       435
Lys Ile Lys Asp Ile Tyr Gly Lys Asp Ala Leu Leu His Glu His Tyr
                130                 135                 140

gtt tac gca aag gag ggc tat gag ccc gtg ctc gtt att cag agt agt       483
Val Tyr Ala Lys Glu Gly Tyr Glu Pro Val Leu Val Ile Gln Ser Ser
            145                 150                 155

gag gac tac gtc gag aat acc gag aaa gct ctg aat gtg tat tac gag       531
Glu Asp Tyr Val Glu Asn Thr Glu Lys Ala Leu Asn Val Tyr Tyr Glu
```

```
        160                 165                 170

atc gga aag att ctg tcc cgg gac atc ctg tcc aaa atc aac cag cca        579
Ile Gly Lys Ile Leu Ser Arg Asp Ile Leu Ser Lys Ile Asn Gln Pro
    175                 180                 185

tac cag aaa ttc ctt gat gtt ctt aac aca atc aaa aac gcg tca gat        627
Tyr Gln Lys Phe Leu Asp Val Leu Asn Thr Ile Lys Asn Ala Ser Asp
190                 195                 200                 205

agc gac ggg cag gat ctt ctg ttt aca aat caa ctc aag gaa cac ccc        675
Ser Asp Gly Gln Asp Leu Leu Phe Thr Asn Gln Leu Lys Glu His Pro
                210                 215                 220

act gat ttc agc gtg gag ttc ctc gag cag aat tct aac gaa gtc cag        723
Thr Asp Phe Ser Val Glu Phe Leu Glu Gln Asn Ser Asn Glu Val Gln
            225                 230                 235

gag gtg ttc gcc aag gca ttt gcg tac tat atc gaa ccc cag cat cgc        771
Glu Val Phe Ala Lys Ala Phe Ala Tyr Tyr Ile Glu Pro Gln His Arg
        240                 245                 250

gat gtg ctc cag ctg tac gcc ccg gag gca ttt aac tac atg gac aaa        819
Asp Val Leu Gln Leu Tyr Ala Pro Glu Ala Phe Asn Tyr Met Asp Lys
    255                 260                 265

ttc aat gaa cag gag att aat ctg tct ctg gag gaa ctg aaa gac cag        867
Phe Asn Glu Gln Glu Ile Asn Leu Ser Leu Glu Glu Leu Lys Asp Gln
270                 275                 280                 285

agg atg ctc tcc cgg tat gaa aag tgg gaa aag atc aaa cag cat tac        915
Arg Met Leu Ser Arg Tyr Glu Lys Trp Glu Lys Ile Lys Gln His Tyr
                290                 295                 300

cag cat tgg tcc gac tcc ctg tca gaa gag ggg cgc ggc ctg ttg aaa        963
Gln His Trp Ser Asp Ser Leu Ser Glu Glu Gly Arg Gly Leu Leu Lys
            305                 310                 315

aag ttg cag att ccc atc gag cct aag aaa gat gat ata ata cac tct       1011
Lys Leu Gln Ile Pro Ile Glu Pro Lys Lys Asp Asp Ile Ile His Ser
        320                 325                 330

cta agc cag gag gag aag gaa ctc ctg aag cgg ata caa atc gac tca       1059
Leu Ser Gln Glu Glu Lys Glu Leu Leu Lys Arg Ile Gln Ile Asp Ser
    335                 340                 345

tcc gat ttc ctt agc aca gaa gag aag gag ttt cta aaa aaa ctt cag       1107
Ser Asp Phe Leu Ser Thr Glu Glu Lys Glu Phe Leu Lys Lys Leu Gln
350                 355                 360                 365

ata gat att aga gat tca ctg agc gag gaa gag aag gag ctg ctc aac       1155
Ile Asp Ile Arg Asp Ser Leu Ser Glu Glu Glu Lys Glu Leu Leu Asn
                370                 375                 380

cga att caa gtc gat agt tcg aac ccc ttg tca gaa aaa gag aag gaa       1203
Arg Ile Gln Val Asp Ser Ser Asn Pro Leu Ser Glu Lys Glu Lys Glu
            385                 390                 395

ttc ctg aaa aag ttg aag ctc gac atc cag ccg tac gat att aat cag       1251
Phe Leu Lys Lys Leu Lys Leu Asp Ile Gln Pro Tyr Asp Ile Asn Gln
        400                 405                 410
```

```
cgg cta caa gac acc ggc ggt ctg att gat agc ccc agc atc aac ctt       1299
Arg Leu Gln Asp Thr Gly Gly Leu Ile Asp Ser Pro Ser Ile Asn Leu
    415                 420                 425

gac gta cgg aag caa tat aag cgc gac att caa aat atc gac gcc cta       1347
Asp Val Arg Lys Gln Tyr Lys Arg Asp Ile Gln Asn Ile Asp Ala Leu
430                 435                 440                 445

tta cat caa tcc ata ggc tcc acg cta tac aat aaa atc tat cta tac       1395
Leu His Gln Ser Ile Gly Ser Thr Leu Tyr Asn Lys Ile Tyr Leu Tyr
                450                 455                 460

gaa aac atg aat att aac aat ctc acc gct aca ctg gga gcg gac ctg       1443
Glu Asn Met Asn Ile Asn Asn Leu Thr Ala Thr Leu Gly Ala Asp Leu
            465                 470                 475

gtc gat agt aca gac aac aca aag ata aac aga ggt att ttc aac gaa       1491
Val Asp Ser Thr Asp Asn Thr Lys Ile Asn Arg Gly Ile Phe Asn Glu
        480                 485                 490

ttc aaa aag aac ttt aag tat tcg atc agc agt aac tat atg att gtt       1539
Phe Lys Lys Asn Phe Lys Tyr Ser Ile Ser Ser Asn Tyr Met Ile Val
    495                 500                 505

gac atc aat gaa cgg ccc gca tta gac aat gag agg ttg aag tgg aga       1587
Asp Ile Asn Glu Arg Pro Ala Leu Asp Asn Glu Arg Leu Lys Trp Arg
510                 515                 520                 525

att caa ctg agt cct gat act agg gcc ggc tat ctg gag aac ggg aaa       1635
Ile Gln Leu Ser Pro Asp Thr Arg Ala Gly Tyr Leu Glu Asn Gly Lys
                530                 535                 540

ctg atc tta cag cga aac atc ggg ctg gag atc aag gat gtg cag att       1683
Leu Ile Leu Gln Arg Asn Ile Gly Leu Glu Ile Lys Asp Val Gln Ile
            545                 550                 555

atc aag cag agc gaa aaa gaa tac att cgc atc gac gcc aag gtg gtg       1731
Ile Lys Gln Ser Glu Lys Glu Tyr Ile Arg Ile Asp Ala Lys Val Val
        560                 565                 570

cct aag tca aag atc gat acc aag atc cag gaa gct cag ctc aac att       1779
Pro Lys Ser Lys Ile Asp Thr Lys Ile Gln Glu Ala Gln Leu Asn Ile
    575                 580                 585

aac cag gag tgg aat aaa gct ctt ggt ctg cca aaa tac acc aaa ctt       1827
Asn Gln Glu Trp Asn Lys Ala Leu Gly Leu Pro Lys Tyr Thr Lys Leu
590                 595                 600                 605

atc acc ttt aat gtg cac aac agg tat gcc tct aat atc gtc gag tca       1875
Ile Thr Phe Asn Val His Asn Arg Tyr Ala Ser Asn Ile Val Glu Ser
                610                 615                 620

gca tac ctg att ctc aat gaa tgg aag aac aat att cag tct gac ctg       1923
Ala Tyr Leu Ile Leu Asn Glu Trp Lys Asn Asn Ile Gln Ser Asp Leu
            625                 630                 635

atc aag aag gtc acg aat tat ctg gtg gac gga aat ggc aga ttc gtg       1971
Ile Lys Lys Val Thr Asn Tyr Leu Val Asp Gly Asn Gly Arg Phe Val
        640                 645                 650

ttc acc gac ata act ttg cca aac att gcc gag caa tac act cat cag       2019
Phe Thr Asp Ile Thr Leu Pro Asn Ile Ala Glu Gln Tyr Thr His Gln
```

```
    655                 660                 665

gat gaa att tac gag caa gtc cac tcc aaa ggt ctg tat gtt cca gag      2067
Asp Glu Ile Tyr Glu Gln Val His Ser Lys Gly Leu Tyr Val Pro Glu
670                 675                 680                 685

tca aga tcg att ctg ctc cat ggt cca tcc aaa ggg gtt gag ctt cga      2115
Ser Arg Ser Ile Leu Leu His Gly Pro Ser Lys Gly Val Glu Leu Arg
                690                 695                 700

aac gat tct gag gga ttt atc gct gac ttt gga gcc gct gtg gat gac      2163
Asn Asp Ser Glu Gly Phe Ile Ala Asp Phe Gly Ala Ala Val Asp Asp
            705                 710                 715

tac gcc gga tac ctg ttg gat aag aat cag tct gat ctc gtg aca aat      2211
Tyr Ala Gly Tyr Leu Leu Asp Lys Asn Gln Ser Asp Leu Val Thr Asn
        720                 725                 730

agc aaa aaa ttc ata gat att ttc aag gag gaa ggg agt aac ctg act      2259
Ser Lys Lys Phe Ile Asp Ile Phe Lys Glu Glu Gly Ser Asn Leu Thr
    735                 740                 745

tcc tat ggc cgc acg aac gag gct gaa ttt ttt gcg gaa gcc ttt aga      2307
Ser Tyr Gly Arg Thr Asn Glu Ala Glu Phe Phe Ala Glu Ala Phe Arg
750                 755                 760                 765

ctt atg cac agc acc gac cat gct gaa agg ttg aag gtg caa aag aat      2355
Leu Met His Ser Thr Asp His Ala Glu Arg Leu Lys Val Gln Lys Asn
                770                 775                 780

gcc cct aaa acc ttc cag ttc ata aat gac cag atc aag ttc atc atc      2403
Ala Pro Lys Thr Phe Gln Phe Ile Asn Asp Gln Ile Lys Phe Ile Ile
            785                 790                 795

aac tct tgaggatcc                                                    2418
Asn Ser


<210>  10

<211>  799

<212>  PRT

<213>  Artificial Sequence


<220>

<223>  Human TPA/synthetic antigen fusion protein

<400>  10

Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
1               5                   10                  15


Ala Val Phe Val Ser Pro Ser Ala Gly Gly His Gly Asp Val Gly Met
            20                  25                  30
```

His Val Lys Glu Lys Glu Lys Asn Lys Asp Glu Asn Lys Arg Lys Asp
35 40 45

Glu Glu Arg Asn Lys Thr Gln Glu Glu His Leu Lys Glu Ile Met Lys
50 55 60

His Ile Val Lys Ile Glu Val Lys Gly Glu Glu Ala Val Lys Lys Glu
65 70 75 80

Ala Ala Glu Lys Leu Leu Glu Lys Val Pro Ser Asp Val Leu Glu Met
85 90 95

Tyr Lys Ala Ile Gly Gly Lys Ile Tyr Ile Val Asp Gly Asp Ile Thr
100 105 110

Lys His Ile Ser Leu Glu Ala Leu Ser Glu Asp Lys Lys Lys Ile Lys
115 120 125

Asp Ile Tyr Gly Lys Asp Ala Leu Leu His Glu His Tyr Val Tyr Ala
130 135 140

Lys Glu Gly Tyr Glu Pro Val Leu Val Ile Gln Ser Ser Glu Asp Tyr
145 150 155 160

Val Glu Asn Thr Glu Lys Ala Leu Asn Val Tyr Tyr Glu Ile Gly Lys
165 170 175

Ile Leu Ser Arg Asp Ile Leu Ser Lys Ile Asn Gln Pro Tyr Gln Lys
180 185 190

Phe Leu Asp Val Leu Asn Thr Ile Lys Asn Ala Ser Asp Ser Asp Gly
195 200 205

Gln Asp Leu Leu Phe Thr Asn Gln Leu Lys Glu His Pro Thr Asp Phe
210 215 220

Ser Val Glu Phe Leu Glu Gln Asn Ser Asn Glu Val Gln Glu Val Phe
225 230 235 240

Ala Lys Ala Phe Ala Tyr Tyr Ile Glu Pro Gln His Arg Asp Val Leu
245 250 255

Gln Leu Tyr Ala Pro Glu Ala Phe Asn Tyr Met Asp Lys Phe Asn Glu
260 265 270

Gln Glu Ile Asn Leu Ser Leu Glu Glu Leu Lys Asp Gln Arg Met Leu
275 280 285

Ser Arg Tyr Glu Lys Trp Glu Lys Ile Lys Gln His Tyr Gln His Trp
290 295 300

Ser Asp Ser Leu Ser Glu Glu Gly Arg Gly Leu Leu Lys Lys Leu Gln
305 310 315 320

Ile Pro Ile Glu Pro Lys Lys Asp Asp Ile Ile His Ser Leu Ser Gln
325 330 335

Glu Glu Lys Glu Leu Leu Lys Arg Ile Gln Ile Asp Ser Ser Asp Phe
340 345 350

Leu Ser Thr Glu Glu Lys Glu Phe Leu Lys Lys Leu Gln Ile Asp Ile
355 360 365

Arg Asp Ser Leu Ser Glu Glu Glu Lys Glu Leu Leu Asn Arg Ile Gln
370 375 380

Val Asp Ser Ser Asn Pro Leu Ser Glu Lys Glu Lys Glu Phe Leu Lys
385 390 395 400

Lys Leu Lys Leu Asp Ile Gln Pro Tyr Asp Ile Asn Gln Arg Leu Gln
405 410 415

Asp Thr Gly Gly Leu Ile Asp Ser Pro Ser Ile Asn Leu Asp Val Arg
420 425 430

Lys Gln Tyr Lys Arg Asp Ile Gln Asn Ile Asp Ala Leu Leu His Gln
435 440 445

Ser Ile Gly Ser Thr Leu Tyr Asn Lys Ile Tyr Leu Tyr Glu Asn Met
450 455 460

Asn Ile Asn Asn Leu Thr Ala Thr Leu Gly Ala Asp Leu Val Asp Ser
465 470 475 480

Thr Asp Asn Thr Lys Ile Asn Arg Gly Ile Phe Asn Glu Phe Lys Lys
485 490 495

Asn Phe Lys Tyr Ser Ile Ser Ser Asn Tyr Met Ile Val Asp Ile Asn
500 505 510

Glu Arg Pro Ala Leu Asp Asn Glu Arg Leu Lys Trp Arg Ile Gln Leu
515 520 525

Ser Pro Asp Thr Arg Ala Gly Tyr Leu Glu Asn Gly Lys Leu Ile Leu

```
    530                 535                 540

Gln Arg Asn Ile Gly Leu Glu Ile Lys Asp Val Gln Ile Ile Lys Gln
545                 550                 555                 560

Ser Glu Lys Glu Tyr Ile Arg Ile Asp Ala Lys Val Val Pro Lys Ser
                565                 570                 575

Lys Ile Asp Thr Lys Ile Gln Glu Ala Gln Leu Asn Ile Asn Gln Glu
            580                 585                 590

Trp Asn Lys Ala Leu Gly Leu Pro Lys Tyr Thr Lys Leu Ile Thr Phe
        595                 600                 605

Asn Val His Asn Arg Tyr Ala Ser Asn Ile Val Glu Ser Ala Tyr Leu
    610                 615                 620

Ile Leu Asn Glu Trp Lys Asn Asn Ile Gln Ser Asp Leu Ile Lys Lys
625                 630                 635                 640

Val Thr Asn Tyr Leu Val Asp Gly Asn Gly Arg Phe Val Phe Thr Asp
                645                 650                 655

Ile Thr Leu Pro Asn Ile Ala Glu Gln Tyr Thr His Gln Asp Glu Ile
            660                 665                 670

Tyr Glu Gln Val His Ser Lys Gly Leu Tyr Val Pro Glu Ser Arg Ser
        675                 680                 685

Ile Leu Leu His Gly Pro Ser Lys Gly Val Glu Leu Arg Asn Asp Ser
    690                 695                 700

Glu Gly Phe Ile Ala Asp Phe Gly Ala Ala Val Asp Asp Tyr Ala Gly
705                 710                 715                 720

Tyr Leu Leu Asp Lys Asn Gln Ser Asp Leu Val Thr Asn Ser Lys Lys
                725                 730                 735

Phe Ile Asp Ile Phe Lys Glu Glu Gly Ser Asn Leu Thr Ser Tyr Gly
            740                 745                 750

Arg Thr Asn Glu Ala Glu Phe Phe Ala Glu Ala Phe Arg Leu Met His
        755                 760                 765

Ser Thr Asp His Ala Glu Arg Leu Lys Val Gln Lys Asn Ala Pro Lys
    770                 775                 780
```

```
Thr Phe Gln Phe Ile Asn Asp Gln Ile Lys Phe Ile Ile Asn Ser
785                 790                 795

<210>  11

<211>  3631

<212>  DNA

<213>  Bacillus anthracis


<220>

<221>  CDS

<222>  (685)..(3111)

<223>


<400>  11
gcatcatatg gcaattatat tcatagcgcg ttttttcccg ttataaaaga taaaaactcg      60

aaaacaggaa aataatatac cttttattct atacagcaac ctaaatatta taatcaactt     120

ttccatagaa attaatcctt ttgttataca tctttattcc cttaacattg tcaaatttca     180

gttattcatt ctggatagtc aataaataga ttacggttat gttagtattt ttttaaaata     240

atagtattaa atagtggaat gcaaatgata aatgggcttt aaacaaaact aatgaaataa     300

tctacaaatg gaatttctcc agttttagat taaaccatac caaaaaaatc acactgtcaa     360

gaaaaatgat agaatcccta cactaattaa cataaccaaa ttggtagtta taggtagaaa     420

cttatttatt tctataatac catgcaaaaa acaaactaaa tattctgttc catactattt     480

tagtaaatta tttagcaagt aaattttggt gtataaacaa agtttatctt aatataaaaa     540

attactttac ttttatacag attaaaatga aaaatttttt atgacaagaa atattgcctt     600

taatttatga ggaaataagt aaaattttct acatacttta ttttattgtt gaaatgttca     660

cttataaaaa aggagagatt aaat atg aat ata aaa aaa gaa ttt ata aaa       711
                           Met Asn Ile Lys Lys Glu Phe Ile Lys
                           1               5

gta att agt atg tca tgt tta gta aca gca att act ttg agt ggt ccc       759
Val Ile Ser Met Ser Cys Leu Val Thr Ala Ile Thr Leu Ser Gly Pro
10                  15                  20                  25

gtc ttt atc ccc ctt gta cag ggg gcg ggc ggt cat ggt gat gta ggt       807
Val Phe Ile Pro Leu Val Gln Gly Ala Gly Gly His Gly Asp Val Gly
                30                  35                  40

atg cac gta aaa gag aaa gag aaa aat aaa gat gag aat aag aga aaa       855
Met His Val Lys Glu Lys Glu Lys Asn Lys Asp Glu Asn Lys Arg Lys
            45                  50                  55
```

```
gat gaa gaa cga aat aaa aca cag gaa gag cat tta aag gaa atc atg      903
Asp Glu Glu Arg Asn Lys Thr Gln Glu Glu His Leu Lys Glu Ile Met
        60                  65                  70

aaa cac att gta aaa ata gaa gta aaa ggg gag gaa gct gtt aaa aaa      951
Lys His Ile Val Lys Ile Glu Val Lys Gly Glu Glu Ala Val Lys Lys
    75                  80                  85

gag gca gca gaa aag cta ctt gag aaa gta cca tct gat gtt tta gag      999
Glu Ala Ala Glu Lys Leu Leu Glu Lys Val Pro Ser Asp Val Leu Glu
90                  95                  100                 105

atg tat aaa gca att gga gga aag ata tat att gtg gat ggt gat att     1047
Met Tyr Lys Ala Ile Gly Gly Lys Ile Tyr Ile Val Asp Gly Asp Ile
                110                 115                 120

aca aaa cat ata tct tta gaa gca tta tct gaa gat aag aaa aaa ata     1095
Thr Lys His Ile Ser Leu Glu Ala Leu Ser Glu Asp Lys Lys Lys Ile
            125                 130                 135

aaa gac att tat ggg aaa gat gct tta tta cat gaa cat tat gta tat     1143
Lys Asp Ile Tyr Gly Lys Asp Ala Leu Leu His Glu His Tyr Val Tyr
        140                 145                 150

gca aaa gaa gga tat gaa ccc gta ctt gta atc caa tct tcg gaa gat     1191
Ala Lys Glu Gly Tyr Glu Pro Val Leu Val Ile Gln Ser Ser Glu Asp
    155                 160                 165

tat gta gaa aat act gaa aag gca ctg aac gtt tat tat gaa ata ggt     1239
Tyr Val Glu Asn Thr Glu Lys Ala Leu Asn Val Tyr Tyr Glu Ile Gly
170                 175                 180                 185

aag ata tta tca agg gat att tta agt aaa att aat caa cca tat cag     1287
Lys Ile Leu Ser Arg Asp Ile Leu Ser Lys Ile Asn Gln Pro Tyr Gln
                190                 195                 200

aaa ttt tta gat gta tta aat acc att aaa aat gca tct gat tca gat     1335
Lys Phe Leu Asp Val Leu Asn Thr Ile Lys Asn Ala Ser Asp Ser Asp
            205                 210                 215

gga caa gat ctt tta ttt act aat cag ctt aag gaa cat ccc aca gac     1383
Gly Gln Asp Leu Leu Phe Thr Asn Gln Leu Lys Glu His Pro Thr Asp
        220                 225                 230

ttt tct gta gaa ttc ttg gaa caa aat agc aat gag gta caa gaa gta     1431
Phe Ser Val Glu Phe Leu Glu Gln Asn Ser Asn Glu Val Gln Glu Val
    235                 240                 245

ttt gcg aaa gct ttt gca tat tat atc gag cca cag cat cgt gat gtt     1479
Phe Ala Lys Ala Phe Ala Tyr Tyr Ile Glu Pro Gln His Arg Asp Val
250                 255                 260                 265

tta cag ctt tat gca ccg gaa gct ttt aat tac atg gat aaa ttt aac     1527
Leu Gln Leu Tyr Ala Pro Glu Ala Phe Asn Tyr Met Asp Lys Phe Asn
                270                 275                 280

gaa caa gaa ata aat cta tcc ttg gaa gaa ctt aaa gat caa cgg atg     1575
Glu Gln Glu Ile Asn Leu Ser Leu Glu Glu Leu Lys Asp Gln Arg Met
            285                 290                 295

ctg tca aga tat gaa aaa tgg gaa aag ata aaa cag cac tat caa cac     1623
Leu Ser Arg Tyr Glu Lys Trp Glu Lys Ile Lys Gln His Tyr Gln His
```

```
        300                 305                 310

tgg agc gat tct tta tct gaa gaa gga aga gga ctt tta aaa aag ctg      1671
Trp Ser Asp Ser Leu Ser Glu Glu Gly Arg Gly Leu Leu Lys Lys Leu
    315                 320                 325

cag att cct att gag cca aag aaa gat gac ata att cat tct tta tct      1719
Gln Ile Pro Ile Glu Pro Lys Lys Asp Asp Ile Ile His Ser Leu Ser
330                 335                 340                 345

caa gaa gaa aaa gag ctt cta aaa aga ata caa att gat agt agt gat      1767
Gln Glu Glu Lys Glu Leu Leu Lys Arg Ile Gln Ile Asp Ser Ser Asp
                350                 355                 360

ttt tta tct act gag gaa aaa gag ttt tta aaa aag cta caa att gat      1815
Phe Leu Ser Thr Glu Glu Lys Glu Phe Leu Lys Lys Leu Gln Ile Asp
            365                 370                 375

att cgt gat tct tta tct gaa gaa gaa aaa gag ctt tta aat aga ata      1863
Ile Arg Asp Ser Leu Ser Glu Glu Glu Lys Glu Leu Leu Asn Arg Ile
        380                 385                 390

cag gtg gat agt agt aat cct tta tct gaa aaa gaa aaa gag ttt tta      1911
Gln Val Asp Ser Ser Asn Pro Leu Ser Glu Lys Glu Lys Glu Phe Leu
    395                 400                 405

aaa aag ctg aaa ctt gat att caa cca tat gat att aat caa agg ttg      1959
Lys Lys Leu Lys Leu Asp Ile Gln Pro Tyr Asp Ile Asn Gln Arg Leu
410                 415                 420                 425

caa gat aca gga ggg tta att gat agt ccg tca att aat ctt gat gta      2007
Gln Asp Thr Gly Gly Leu Ile Asp Ser Pro Ser Ile Asn Leu Asp Val
                430                 435                 440

aga aag cag tat aaa agg gat att caa aat att gat gct tta tta cat      2055
Arg Lys Gln Tyr Lys Arg Asp Ile Gln Asn Ile Asp Ala Leu Leu His
            445                 450                 455

caa tcc att gga agt acc ttg tac aat aaa att tat ttg tat gaa aat      2103
Gln Ser Ile Gly Ser Thr Leu Tyr Asn Lys Ile Tyr Leu Tyr Glu Asn
        460                 465                 470

atg aat atc aat aac ctt aca gca acc cta ggt gcg gat tta gtt gat      2151
Met Asn Ile Asn Asn Leu Thr Ala Thr Leu Gly Ala Asp Leu Val Asp
    475                 480                 485

tcc act gat aat act aaa att aat aga ggt att ttc aat gaa ttc aaa      2199
Ser Thr Asp Asn Thr Lys Ile Asn Arg Gly Ile Phe Asn Glu Phe Lys
490                 495                 500                 505

aaa aat ttc aaa tat agt att tct agt aac tat atg att gtt gat ata      2247
Lys Asn Phe Lys Tyr Ser Ile Ser Ser Asn Tyr Met Ile Val Asp Ile
                510                 515                 520

aat gaa agg cct gca tta gat aat gag cgt ttg aaa tgg aga atc caa      2295
Asn Glu Arg Pro Ala Leu Asp Asn Glu Arg Leu Lys Trp Arg Ile Gln
            525                 530                 535

tta tca cca gat act cga gca gga tat tta gaa aat gga aag ctt ata      2343
Leu Ser Pro Asp Thr Arg Ala Gly Tyr Leu Glu Asn Gly Lys Leu Ile
        540                 545                 550
```

```
tta caa aga aac atc ggt ctg gaa ata aag gat gta caa ata att aag      2391
Leu Gln Arg Asn Ile Gly Leu Glu Ile Lys Asp Val Gln Ile Ile Lys
    555                 560                 565

caa tcc gaa aaa gaa tat ata agg att gat gcg aaa gta gtg cca aag      2439
Gln Ser Glu Lys Glu Tyr Ile Arg Ile Asp Ala Lys Val Val Pro Lys
570                 575                 580                 585

agt aaa ata gat aca aaa att caa gaa gca cag tta aat ata aat cag      2487
Ser Lys Ile Asp Thr Lys Ile Gln Glu Ala Gln Leu Asn Ile Asn Gln
                590                 595                 600

gaa tgg aat aaa gca tta ggg tta cca aaa tat aca aag ctt att aca      2535
Glu Trp Asn Lys Ala Leu Gly Leu Pro Lys Tyr Thr Lys Leu Ile Thr
            605                 610                 615

ttc aac gtg cat aat aga tat gca tcc aat att gta gaa agt gct tat      2583
Phe Asn Val His Asn Arg Tyr Ala Ser Asn Ile Val Glu Ser Ala Tyr
        620                 625                 630

tta ata ttg aat gaa tgg aaa aat aat att caa agt gat ctt ata aaa      2631
Leu Ile Leu Asn Glu Trp Lys Asn Asn Ile Gln Ser Asp Leu Ile Lys
    635                 640                 645

aag gta aca aat tac tta gtt gat ggt aat gga aga ttt gtt ttt acc      2679
Lys Val Thr Asn Tyr Leu Val Asp Gly Asn Gly Arg Phe Val Phe Thr
650                 655                 660                 665

gat att act ctc cct aat ata gct gaa caa tat aca cat caa gat gag      2727
Asp Ile Thr Leu Pro Asn Ile Ala Glu Gln Tyr Thr His Gln Asp Glu
                670                 675                 680

ata tat gag caa gtt cat tca aaa ggg tta tat gtt cca gaa tcc cgt      2775
Ile Tyr Glu Gln Val His Ser Lys Gly Leu Tyr Val Pro Glu Ser Arg
            685                 690                 695

tct ata tta ctc cat gga cct tca aaa ggt gta gaa tta agg aat gat      2823
Ser Ile Leu Leu His Gly Pro Ser Lys Gly Val Glu Leu Arg Asn Asp
        700                 705                 710

agt gag ggt ttt ata cac gaa ttt gga cat gct gtg gat gat tat gct      2871
Ser Glu Gly Phe Ile His Glu Phe Gly His Ala Val Asp Asp Tyr Ala
    715                 720                 725

gga tat cta tta gat aag aac caa tct gat tta gtt aca aat tct aaa      2919
Gly Tyr Leu Leu Asp Lys Asn Gln Ser Asp Leu Val Thr Asn Ser Lys
730                 735                 740                 745

aaa ttc att gat att ttt aag gaa gaa ggg agt aat tta act tcg tat      2967
Lys Phe Ile Asp Ile Phe Lys Glu Glu Gly Ser Asn Leu Thr Ser Tyr
                750                 755                 760

ggg aga aca aat gaa gcg gaa ttt ttt gca gaa gcc ttt agg tta atg      3015
Gly Arg Thr Asn Glu Ala Glu Phe Phe Ala Glu Ala Phe Arg Leu Met
            765                 770                 775

cat tct acg gac cat gct gaa cgt tta aaa gtt caa aaa aat gct ccg      3063
His Ser Thr Asp His Ala Glu Arg Leu Lys Val Gln Lys Asn Ala Pro
        780                 785                 790

aaa act ttc caa ttt att aac gat cag att aag ttc att att aac tca      3111
Lys Thr Phe Gln Phe Ile Asn Asp Gln Ile Lys Phe Ile Ile Asn Ser
```

795 800 805

taagtaatgt attaaaaatt ttcaaatgga tttaataata ataataataa taataataac 3171

gggaccagcc attatgaagc aactaattct agacttgata gtaattcttg ggaagcacca 3231

gatagtgtaa aaggtggcat tgccagaatg atattttatg tgttcgttag atatgaaggc 3291

aaaaacaatg atcctgacct agaacttaat gataatgtta ttaataattt aatgcctttt 3351

ataggaatat tagtaaaagt gccgaaaaga tcctgttgca aagcttttaa agaacatatt 3411

attctatcaa gtggctgtat attttgttaa ttttcaataa attttgtaat taagcatacg 3471

tcaaaaaacc gaaatctgag ctcagatttc ggtttttga cgtatcttat acagatttgc 3531

aggttttaat ttcatttgct caatattaca aggatcagtt taaatgcaag tatgatgcac 3591

atacatgttg ttaaagacaa taagacgaaa aatatttctc 3631

<210> 12

<211> 809

<212> PRT

<213> Bacillus anthracis

<400> 12

Met Asn Ile Lys Lys Glu Phe Ile Lys Val Ile Ser Met Ser Cys Leu
1 5 10 15

Val Thr Ala Ile Thr Leu Ser Gly Pro Val Phe Ile Pro Leu Val Gln
20 25 30

Gly Ala Gly Gly His Gly Asp Val Gly Met His Val Lys Glu Lys Glu
35 40 45

Lys Asn Lys Asp Glu Asn Lys Arg Lys Asp Glu Glu Arg Asn Lys Thr
50 55 60

Gln Glu Glu His Leu Lys Glu Ile Met Lys His Ile Val Lys Ile Glu
65 70 75 80

Val Lys Gly Glu Glu Ala Val Lys Lys Glu Ala Ala Glu Lys Leu Leu
85 90 95

Glu Lys Val Pro Ser Asp Val Leu Glu Met Tyr Lys Ala Ile Gly Gly
100 105 110

Lys Ile Tyr Ile Val Asp Gly Asp Ile Thr Lys His Ile Ser Leu Glu

115 120 125

Ala Leu Ser Glu Asp Lys Lys Lys Ile Lys Asp Ile Tyr Gly Lys Asp
130 135 140

Ala Leu Leu His Glu His Tyr Val Tyr Ala Lys Glu Gly Tyr Glu Pro
145 150 155 160

Val Leu Val Ile Gln Ser Ser Glu Asp Tyr Val Glu Asn Thr Glu Lys
165 170 175

Ala Leu Asn Val Tyr Tyr Glu Ile Gly Lys Ile Leu Ser Arg Asp Ile
180 185 190

Leu Ser Lys Ile Asn Gln Pro Tyr Gln Lys Phe Leu Asp Val Leu Asn
195 200 205

Thr Ile Lys Asn Ala Ser Asp Ser Asp Gly Gln Asp Leu Leu Phe Thr
210 215 220

Asn Gln Leu Lys Glu His Pro Thr Asp Phe Ser Val Glu Phe Leu Glu
225 230 235 240

Gln Asn Ser Asn Glu Val Gln Glu Val Phe Ala Lys Ala Phe Ala Tyr
245 250 255

Tyr Ile Glu Pro Gln His Arg Asp Val Leu Gln Leu Tyr Ala Pro Glu
260 265 270

Ala Phe Asn Tyr Met Asp Lys Phe Asn Glu Gln Glu Ile Asn Leu Ser
275 280 285

Leu Glu Glu Leu Lys Asp Gln Arg Met Leu Ser Arg Tyr Glu Lys Trp
290 295 300

Glu Lys Ile Lys Gln His Tyr Gln His Trp Ser Asp Ser Leu Ser Glu
305 310 315 320

Glu Gly Arg Gly Leu Leu Lys Lys Leu Gln Ile Pro Ile Glu Pro Lys
325 330 335

Lys Asp Asp Ile Ile His Ser Leu Ser Gln Glu Glu Lys Glu Leu Leu
340 345 350

Lys Arg Ile Gln Ile Asp Ser Ser Asp Phe Leu Ser Thr Glu Glu Lys
355 360 365

Glu Phe Leu Lys Lys Leu Gln Ile Asp Ile Arg Asp Ser Leu Ser Glu
370 375 380

Glu Glu Lys Glu Leu Leu Asn Arg Ile Gln Val Asp Ser Ser Asn Pro
385 390 395 400

Leu Ser Glu Lys Glu Lys Glu Phe Leu Lys Lys Leu Lys Leu Asp Ile
405 410 415

Gln Pro Tyr Asp Ile Asn Gln Arg Leu Gln Asp Thr Gly Gly Leu Ile
420 425 430

Asp Ser Pro Ser Ile Asn Leu Asp Val Arg Lys Gln Tyr Lys Arg Asp
435 440 445

Ile Gln Asn Ile Asp Ala Leu Leu His Gln Ser Ile Gly Ser Thr Leu
450 455 460

Tyr Asn Lys Ile Tyr Leu Tyr Glu Asn Met Asn Ile Asn Asn Leu Thr
465 470 475 480

Ala Thr Leu Gly Ala Asp Leu Val Asp Ser Thr Asp Asn Thr Lys Ile
485 490 495

Asn Arg Gly Ile Phe Asn Glu Phe Lys Lys Asn Phe Lys Tyr Ser Ile
500 505 510

Ser Ser Asn Tyr Met Ile Val Asp Ile Asn Glu Arg Pro Ala Leu Asp
515 520 525

Asn Glu Arg Leu Lys Trp Arg Ile Gln Leu Ser Pro Asp Thr Arg Ala
530 535 540

Gly Tyr Leu Glu Asn Gly Lys Leu Ile Leu Gln Arg Asn Ile Gly Leu
545 550 555 560

Glu Ile Lys Asp Val Gln Ile Ile Lys Gln Ser Glu Lys Glu Tyr Ile
565 570 575

Arg Ile Asp Ala Lys Val Val Pro Lys Ser Lys Ile Asp Thr Lys Ile
580 585 590

Gln Glu Ala Gln Leu Asn Ile Asn Gln Glu Trp Asn Lys Ala Leu Gly
595 600 605

Leu Pro Lys Tyr Thr Lys Leu Ile Thr Phe Asn Val His Asn Arg Tyr
610 615 620

Ala Ser Asn Ile Val Glu Ser Ala Tyr Leu Ile Leu Asn Glu Trp Lys
625 630 635 640

Asn Asn Ile Gln Ser Asp Leu Ile Lys Lys Val Thr Asn Tyr Leu Val
645 650 655

Asp Gly Asn Gly Arg Phe Val Phe Thr Asp Ile Thr Leu Pro Asn Ile
660 665 670

Ala Glu Gln Tyr Thr His Gln Asp Glu Ile Tyr Glu Gln Val His Ser
675 680 685

Lys Gly Leu Tyr Val Pro Glu Ser Arg Ser Ile Leu Leu His Gly Pro
690 695 700

Ser Lys Gly Val Glu Leu Arg Asn Asp Ser Glu Gly Phe Ile His Glu
705 710 715 720

Phe Gly His Ala Val Asp Asp Tyr Ala Gly Tyr Leu Leu Asp Lys Asn
725 730 735

Gln Ser Asp Leu Val Thr Asn Ser Lys Lys Phe Ile Asp Ile Phe Lys
740 745 750

Glu Glu Gly Ser Asn Leu Thr Ser Tyr Gly Arg Thr Asn Glu Ala Glu
755 760 765

Phe Phe Ala Glu Ala Phe Arg Leu Met His Ser Thr Asp His Ala Glu
770 775 780

Arg Leu Lys Val Gln Lys Asn Ala Pro Lys Thr Phe Gln Phe Ile Asn
785 790 795 800

Asp Gln Ile Lys Phe Ile Ile Asn Ser
805

<210> 13

<211> 1740

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic coding region for Human TPA/B. anthracis

antigen fusion protein

<220>

<221> CDS

<222> (13)..(1731)

<223>

<400> 13

gatatcgcca cc atg gat gca atg aag aga ggg ctc tgc tgt gtg ctg ctg 51
Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu
1 5 10

ctg tgt gga gca gtc ttc gtt tcg ccc agc gcc ggc ggg cat ggg gac 99
Leu Cys Gly Ala Val Phe Val Ser Pro Ser Ala Gly Gly His Gly Asp
15 20 25

gtt ggc atg cat gtg aaa gaa aag gag aaa aac aag gac gaa aac aag 147
Val Gly Met His Val Lys Glu Lys Glu Lys Asn Lys Asp Glu Asn Lys
30 35 40 45

cgt aaa gac gaa gaa cgt aat aaa aca cag gag gaa cac tta aag gag 195
Arg Lys Asp Glu Glu Arg Asn Lys Thr Gln Glu Glu His Leu Lys Glu
50 55 60

atc atg aag cac ata gta aag att gag gta aaa ggc gaa gag gct gta 243
Ile Met Lys His Ile Val Lys Ile Glu Val Lys Gly Glu Glu Ala Val
65 70 75

aag aag gag gca gca gaa aaa ctg ttg gag aag gtg cct tct gac gtc 291
Lys Lys Glu Ala Ala Glu Lys Leu Leu Glu Lys Val Pro Ser Asp Val
80 85 90

tta gag atg tat aag gcc atc ggc ggt aag atc tat atc gtg gac gga 339
Leu Glu Met Tyr Lys Ala Ile Gly Gly Lys Ile Tyr Ile Val Asp Gly
95 100 105

gac atc act aaa cac ata tct ctc gaa gct ctc tcc gag gac aag aaa 387
Asp Ile Thr Lys His Ile Ser Leu Glu Ala Leu Ser Glu Asp Lys Lys
110 115 120 125

aag att aaa gac atc tac ggg aag gat gcc tta ttg cac gag cac tac 435
Lys Ile Lys Asp Ile Tyr Gly Lys Asp Ala Leu Leu His Glu His Tyr
130 135 140

gtt tac gca aag gag ggc tat gag ccc gtg ctc gtt att cag agt agt 483
Val Tyr Ala Lys Glu Gly Tyr Glu Pro Val Leu Val Ile Gln Ser Ser
145 150 155

gag gac tac gtc gag aat acc gag aaa gct ctg aat gtg tat tac gag 531
Glu Asp Tyr Val Glu Asn Thr Glu Lys Ala Leu Asn Val Tyr Tyr Glu
160 165 170

atc gga aag att ctg tcc cgg gac atc ctg tcc aaa atc aac cag cca 579
Ile Gly Lys Ile Leu Ser Arg Asp Ile Leu Ser Lys Ile Asn Gln Pro
175 180 185

tac cag aaa ttc ctt gat gtt ctt aac aca atc aaa aac gcg tca gat 627

```
Tyr Gln Lys Phe Leu Asp Val Leu Asn Thr Ile Lys Asn Ala Ser Asp
190                 195                 200                 205

agc gac ggg cag gat ctt ctg ttt aca aat caa ctc aag gaa cac ccc      675
Ser Asp Gly Gln Asp Leu Leu Phe Thr Asn Gln Leu Lys Glu His Pro
                210                 215                 220

act gat ttc agc gtg gag ttc ctc gag cag aat tct aac gaa gtc cag      723
Thr Asp Phe Ser Val Glu Phe Leu Glu Gln Asn Ser Asn Glu Val Gln
            225                 230                 235

gag gtg ttc gcc aag gca ttt gcg tac tat atc gaa ccc cag cat cgc      771
Glu Val Phe Ala Lys Ala Phe Ala Tyr Tyr Ile Glu Pro Gln His Arg
        240                 245                 250

gat gtg ctc cag ctg tac gcc ccg gag gca ttt aac tac atg gac aaa      819
Asp Val Leu Gln Leu Tyr Ala Pro Glu Ala Phe Asn Tyr Met Asp Lys
    255                 260                 265

ttc aat gaa cag gag att aat ctg tct ctg gag gaa ctg aaa gac cag      867
Phe Asn Glu Gln Glu Ile Asn Leu Ser Leu Glu Glu Leu Lys Asp Gln
270                 275                 280                 285

agg atg ctc tcc cgg tat gaa aag tgg gaa aag atc aaa cag cat tac      915
Arg Met Leu Ser Arg Tyr Glu Lys Trp Glu Lys Ile Lys Gln His Tyr
                290                 295                 300

cag cat tgg tcc gac tcc ctg tca gaa gag ggg cgc ggc ctg ttg aaa      963
Gln His Trp Ser Asp Ser Leu Ser Glu Glu Gly Arg Gly Leu Leu Lys
            305                 310                 315

aag ttg cag att ccc atc gag cct aag aaa gat gat ata ata cac tct     1011
Lys Leu Gln Ile Pro Ile Glu Pro Lys Lys Asp Asp Ile Ile His Ser
        320                 325                 330

cta agc cag gag gag aag gaa ctc ctg aag cgg ata caa atc gac tca     1059
Leu Ser Gln Glu Glu Lys Glu Leu Leu Lys Arg Ile Gln Ile Asp Ser
    335                 340                 345

tcc gat ttc ctt agc aca gaa gag aag gag ttt cta aaa aaa ctt cag     1107
Ser Asp Phe Leu Ser Thr Glu Glu Lys Glu Phe Leu Lys Lys Leu Gln
350                 355                 360                 365

ata gat att aga gat tca ctg agc gag gaa gag aag gag ctg ctc aac     1155
Ile Asp Ile Arg Asp Ser Leu Ser Glu Glu Glu Lys Glu Leu Leu Asn
                370                 375                 380

cga att caa gtc gat agt tcg aac ccc ttg tca gaa aaa gag aag gaa     1203
Arg Ile Gln Val Asp Ser Ser Asn Pro Leu Ser Glu Lys Glu Lys Glu
            385                 390                 395

ttc ctg aaa aag ttg aag ctc gac atc cag ccg tac gat att aat cag     1251
Phe Leu Lys Lys Leu Lys Leu Asp Ile Gln Pro Tyr Asp Ile Asn Gln
        400                 405                 410

cgg cta caa gac acc ggc ggt ctg att gat agc ccc agc atc aac ctt     1299
Arg Leu Gln Asp Thr Gly Gly Leu Ile Asp Ser Pro Ser Ile Asn Leu
    415                 420                 425

gac gta cgg aag caa tat aag cgc gac att caa aat atc gac gcc cta     1347
Asp Val Arg Lys Gln Tyr Lys Arg Asp Ile Gln Asn Ile Asp Ala Leu
430                 435                 440                 445
```

```
tta cat caa tcc ata ggc tcc acg cta tac aat aaa atc tat cta tac      1395
Leu His Gln Ser Ile Gly Ser Thr Leu Tyr Asn Lys Ile Tyr Leu Tyr
                450                 455                 460

gaa aac atg aat att aac aat ctc acc gct aca ctg gga gcg gac ctg      1443
Glu Asn Met Asn Ile Asn Asn Leu Thr Ala Thr Leu Gly Ala Asp Leu
            465                 470                 475

gtc gat agt aca gac aac aca aag ata aac aga ggt att ttc aac gaa      1491
Val Asp Ser Thr Asp Asn Thr Lys Ile Asn Arg Gly Ile Phe Asn Glu
        480                 485                 490

ttc aaa aag aac ttt aag tat tcg atc agc agt aac tat atg att gtt      1539
Phe Lys Lys Asn Phe Lys Tyr Ser Ile Ser Ser Asn Tyr Met Ile Val
    495                 500                 505

gac atc aat gaa cgg ccc gca tta gac aat gag agg ttg aag tgg aga      1587
Asp Ile Asn Glu Arg Pro Ala Leu Asp Asn Glu Arg Leu Lys Trp Arg
510                 515                 520                 525

att caa ctg agt cct gat act agg gcc ggc tat ctg gag aac ggg aaa      1635
Ile Gln Leu Ser Pro Asp Thr Arg Ala Gly Tyr Leu Glu Asn Gly Lys
                530                 535                 540

ctg atc tta cag cga aac atc ggg ctg gag atc aag gat gtg cag att      1683
Leu Ile Leu Gln Arg Asn Ile Gly Leu Glu Ile Lys Asp Val Gln Ile
            545                 550                 555

atc aag cag agc gaa aaa gaa tac att cgc atc gac gcc aag gtg gtg      1731
Ile Lys Gln Ser Glu Lys Glu Tyr Ile Arg Ile Asp Ala Lys Val Val
        560                 565                 570

tagggatcc                                                            1740
```

<210> 14

<211> 573

<212> PRT

<213> Artificial Sequence

<220>

<223> Human TPA/B. anthracis antigen fusion protein

<400> 14

```
Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
1               5                   10                  15

Ala Val Phe Val Ser Pro Ser Ala Gly Gly His Gly Asp Val Gly Met
            20                  25                  30

His Val Lys Glu Lys Glu Lys Asn Lys Asp Glu Asn Lys Arg Lys Asp
        35                  40                  45
```

```
Glu Glu Arg Asn Lys Thr Gln Glu Glu His Leu Lys Glu Ile Met Lys
    50                  55                  60

His Ile Val Lys Ile Glu Val Lys Gly Glu Glu Ala Val Lys Lys Glu
65                  70                  75                  80

Ala Ala Glu Lys Leu Leu Glu Lys Val Pro Ser Asp Val Leu Glu Met
                85                  90                  95

Tyr Lys Ala Ile Gly Gly Lys Ile Tyr Ile Val Asp Gly Asp Ile Thr
            100                 105                 110

Lys His Ile Ser Leu Glu Ala Leu Ser Glu Asp Lys Lys Lys Ile Lys
        115                 120                 125

Asp Ile Tyr Gly Lys Asp Ala Leu Leu His Glu His Tyr Val Tyr Ala
    130                 135                 140

Lys Glu Gly Tyr Glu Pro Val Leu Val Ile Gln Ser Ser Glu Asp Tyr
145                 150                 155                 160

Val Glu Asn Thr Glu Lys Ala Leu Asn Val Tyr Tyr Glu Ile Gly Lys
                165                 170                 175

Ile Leu Ser Arg Asp Ile Leu Ser Lys Ile Asn Gln Pro Tyr Gln Lys
            180                 185                 190

Phe Leu Asp Val Leu Asn Thr Ile Lys Asn Ala Ser Asp Ser Asp Gly
        195                 200                 205

Gln Asp Leu Leu Phe Thr Asn Gln Leu Lys Glu His Pro Thr Asp Phe
    210                 215                 220

Ser Val Glu Phe Leu Glu Gln Asn Ser Asn Glu Val Gln Glu Val Phe
225                 230                 235                 240

Ala Lys Ala Phe Ala Tyr Tyr Ile Glu Pro Gln His Arg Asp Val Leu
                245                 250                 255

Gln Leu Tyr Ala Pro Glu Ala Phe Asn Tyr Met Asp Lys Phe Asn Glu
            260                 265                 270

Gln Glu Ile Asn Leu Ser Leu Glu Glu Leu Lys Asp Gln Arg Met Leu
        275                 280                 285

Ser Arg Tyr Glu Lys Trp Glu Lys Ile Lys Gln His Tyr Gln His Trp
    290                 295                 300
```

```
Ser Asp Ser Leu Ser Glu Glu Gly Arg Gly Leu Leu Lys Lys Leu Gln
305                 310                 315                 320

Ile Pro Ile Glu Pro Lys Lys Asp Asp Ile Ile His Ser Leu Ser Gln
                325                 330                 335

Glu Glu Lys Glu Leu Leu Lys Arg Ile Gln Ile Asp Ser Ser Asp Phe
            340                 345                 350

Leu Ser Thr Glu Glu Lys Glu Phe Leu Lys Lys Leu Gln Ile Asp Ile
        355                 360                 365

Arg Asp Ser Leu Ser Glu Glu Glu Lys Glu Leu Leu Asn Arg Ile Gln
    370                 375                 380

Val Asp Ser Ser Asn Pro Leu Ser Glu Lys Glu Lys Glu Phe Leu Lys
385                 390                 395                 400

Lys Leu Lys Leu Asp Ile Gln Pro Tyr Asp Ile Asn Gln Arg Leu Gln
                405                 410                 415

Asp Thr Gly Gly Leu Ile Asp Ser Pro Ser Ile Asn Leu Asp Val Arg
            420                 425                 430

Lys Gln Tyr Lys Arg Asp Ile Gln Asn Ile Asp Ala Leu Leu His Gln
        435                 440                 445

Ser Ile Gly Ser Thr Leu Tyr Asn Lys Ile Tyr Leu Tyr Glu Asn Met
    450                 455                 460

Asn Ile Asn Asn Leu Thr Ala Thr Leu Gly Ala Asp Leu Val Asp Ser
465                 470                 475                 480

Thr Asp Asn Thr Lys Ile Asn Arg Gly Ile Phe Asn Glu Phe Lys Lys
                485                 490                 495

Asn Phe Lys Tyr Ser Ile Ser Ser Asn Tyr Met Ile Val Asp Ile Asn
            500                 505                 510

Glu Arg Pro Ala Leu Asp Asn Glu Arg Leu Lys Trp Arg Ile Gln Leu
        515                 520                 525

Ser Pro Asp Thr Arg Ala Gly Tyr Leu Glu Asn Gly Lys Leu Ile Leu
    530                 535                 540

Gln Arg Asn Ile Gly Leu Glu Ile Lys Asp Val Gln Ile Ile Lys Gln
```

545 550 555 560

Ser Glu Lys Glu Tyr Ile Arg Ile Asp Ala Lys Val Val
565 570

<210> 15

<211> 753

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic coding region for Human TPA/B. anthracis antigen fusion protein

<220>

<221> CDS

<222> (13)..(744)

<223>

<400> 15

```
gatatcgcca cc atg gat gca atg aag aga ggg ctc tgc tgt gtg ctg ctg      51
              Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu
              1               5                   10

ctg tgt gga gca gtc ttc gtt tcg ccc agc gcc ggc ggg cat ggg gac        99
Leu Cys Gly Ala Val Phe Val Ser Pro Ser Ala Gly Gly His Gly Asp
    15                  20                  25

gtt ggc atg cat gtg aaa gaa aag gag aaa aac aag gac gaa aac aag       147
Val Gly Met His Val Lys Glu Lys Glu Lys Asn Lys Asp Glu Asn Lys
30                  35                  40                  45

cgt aaa gac gaa gaa cgt aat aaa aca cag gag gaa cac tta aag gag       195
Arg Lys Asp Glu Glu Arg Asn Lys Thr Gln Glu Glu His Leu Lys Glu
                50                  55                  60

atc atg aag cac ata gta aag att gag gta aaa ggc gaa gag gct gta       243
Ile Met Lys His Ile Val Lys Ile Glu Val Lys Gly Glu Glu Ala Val
            65                  70                  75

aag aag gag gca gca gaa aaa ctg ttg gag aag gtg cct tct gac gtc       291
Lys Lys Glu Ala Ala Glu Lys Leu Leu Glu Lys Val Pro Ser Asp Val
        80                  85                  90

tta gag atg tat aag gcc atc ggc ggt aag atc tat atc gtg gac gga       339
Leu Glu Met Tyr Lys Ala Ile Gly Gly Lys Ile Tyr Ile Val Asp Gly
    95                  100                 105

gac atc act aaa cac ata tct ctc gaa gct ctc tcc gag gac aag aaa       387
```

```
Asp Ile Thr Lys His Ile Ser Leu Glu Ala Leu Ser Glu Asp Lys Lys
110                 115                 120                 125

aag att aaa gac atc tac ggg aag gat gcc tta ttg cac gag cac tac      435
Lys Ile Lys Asp Ile Tyr Gly Lys Asp Ala Leu Leu His Glu His Tyr
                130                 135                 140

gtt tac gca aag gag ggc tat gag ccc gtg ctc gtt att cag agt agt      483
Val Tyr Ala Lys Glu Gly Tyr Glu Pro Val Leu Val Ile Gln Ser Ser
            145                 150                 155

gag gac tac gtc gag aat acc gag aaa gct ctg aat gtg tat tac gag      531
Glu Asp Tyr Val Glu Asn Thr Glu Lys Ala Leu Asn Val Tyr Tyr Glu
        160                 165                 170

atc gga aag att ctg tcc cgg gac atc ctg tcc aaa atc aac cag cca      579
Ile Gly Lys Ile Leu Ser Arg Asp Ile Leu Ser Lys Ile Asn Gln Pro
    175                 180                 185

tac cag aaa ttc ctt gat gtt ctt aac aca atc aaa aac gcg tca gat      627
Tyr Gln Lys Phe Leu Asp Val Leu Asn Thr Ile Lys Asn Ala Ser Asp
190                 195                 200                 205

agc gac ggg cag gat ctt ctg ttt aca aat caa ctc aag gaa cac ccc      675
Ser Asp Gly Gln Asp Leu Leu Phe Thr Asn Gln Leu Lys Glu His Pro
                210                 215                 220

act gat ttc agc gtg gag ttc ctc gag cag aat tct aac gaa gtc cag      723
Thr Asp Phe Ser Val Glu Phe Leu Glu Gln Asn Ser Asn Glu Val Gln
            225                 230                 235

gag gtg ttc gcc aag gca ttt tgaggatcc                                753
Glu Val Phe Ala Lys Ala Phe
        240
```

<210> 16

<211> 244

<212> PRT

<213> Artificial Sequence

<220>

<223> Human TPA/B. anthracis antigen fusion protein

<400> 16

```
Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
1               5                   10                  15

Ala Val Phe Val Ser Pro Ser Ala Gly Gly His Gly Asp Val Gly Met
            20                  25                  30

His Val Lys Glu Lys Glu Lys Asn Lys Asp Glu Asn Lys Arg Lys Asp
        35                  40                  45
```

```
Glu Glu Arg Asn Lys Thr Gln Glu Glu His Leu Lys Glu Ile Met Lys
    50                  55                  60

His Ile Val Lys Ile Glu Val Lys Gly Glu Glu Ala Val Lys Lys Glu
65                  70                  75                  80

Ala Ala Glu Lys Leu Leu Glu Lys Val Pro Ser Asp Val Leu Glu Met
                85                  90                  95

Tyr Lys Ala Ile Gly Gly Lys Ile Tyr Ile Val Asp Gly Asp Ile Thr
            100                 105                 110

Lys His Ile Ser Leu Glu Ala Leu Ser Glu Asp Lys Lys Lys Ile Lys
        115                 120                 125

Asp Ile Tyr Gly Lys Asp Ala Leu Leu His Glu His Tyr Val Tyr Ala
    130                 135                 140

Lys Glu Gly Tyr Glu Pro Val Leu Val Ile Gln Ser Ser Glu Asp Tyr
145                 150                 155                 160

Val Glu Asn Thr Glu Lys Ala Leu Asn Val Tyr Tyr Glu Ile Gly Lys
                165                 170                 175

Ile Leu Ser Arg Asp Ile Leu Ser Lys Ile Asn Gln Pro Tyr Gln Lys
            180                 185                 190

Phe Leu Asp Val Leu Asn Thr Ile Lys Asn Ala Ser Asp Ser Asp Gly
        195                 200                 205

Gln Asp Leu Leu Phe Thr Asn Gln Leu Lys Glu His Pro Thr Asp Phe
    210                 215                 220

Ser Val Glu Phe Leu Glu Gln Asn Ser Asn Glu Val Gln Glu Val Phe
225                 230                 235                 240

Ala Lys Ala Phe
```

<210> 17

<211> 1788

<212> DNA

<213> Artificial Sequence

```
<220>

<223>  Synthetic coding region for Human TPA/synthetic
       antigen fusion protein

<220>

<221>  CDS

<222>  (13)..(1779)

<223>


<400>  17
gatatcgcca cc atg gat gca atg aag aga ggg ctc tgc tgt gtg ctg ctg       51
              Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu
              1               5                   10

ctg tgt gga gca gtc ttc gtt tcg ccc agc agc gct ggg cca act gtg         99
Leu Cys Gly Ala Val Phe Val Ser Pro Ser Ser Ala Gly Pro Thr Val
    15                  20                  25

ccc gac aga gac aat gat gga atc cct gat agt cta gag gtt gag gga        147
Pro Asp Arg Asp Asn Asp Gly Ile Pro Asp Ser Leu Glu Val Glu Gly
30                  35                  40                  45

tac acg gta gat gtc aag aac aaa agg act ttt ctc tcg cct tgg atc        195
Tyr Thr Val Asp Val Lys Asn Lys Arg Thr Phe Leu Ser Pro Trp Ile
                50                  55                  60

tca aat atc cat gag aag aag ggg ctt acc aag tac aag tcc tcc ccc        243
Ser Asn Ile His Glu Lys Lys Gly Leu Thr Lys Tyr Lys Ser Ser Pro
            65                  70                  75

gag aag tgg tct acc gct tcc gat cca tat agc gat ttc gag aag gtc        291
Glu Lys Trp Ser Thr Ala Ser Asp Pro Tyr Ser Asp Phe Glu Lys Val
        80                  85                  90

aca ggc cgg atc gat aaa cag gtg tct cca gag gct aga cac ccc ctg        339
Thr Gly Arg Ile Asp Lys Gln Val Ser Pro Glu Ala Arg His Pro Leu
    95                  100                 105

gta gca gcc tac ccg att gta cac gtg gac atg gag aac atc att cta        387
Val Ala Ala Tyr Pro Ile Val His Val Asp Met Glu Asn Ile Ile Leu
110                 115                 120                 125

agc aaa aac gag gac cag tcc aca caa aac act gac tcc gag acc cgc        435
Ser Lys Asn Glu Asp Gln Ser Thr Gln Asn Thr Asp Ser Glu Thr Arg
                130                 135                 140

acc ata tct aaa cag acc agt act tca agg acc cac acc tct gaa gtg        483
Thr Ile Ser Lys Gln Thr Ser Thr Ser Arg Thr His Thr Ser Glu Val
            145                 150                 155

cac ggc aat gcg gaa gtc cat gca tcg ttt ttc gat att ggt ggc tcc        531
His Gly Asn Ala Glu Val His Ala Ser Phe Phe Asp Ile Gly Gly Ser
        160                 165                 170

gtg tca gcc ggc ttt agc aat agc cag tcc tcg acg gtt gcc att gac        579
Val Ser Ala Gly Phe Ser Asn Ser Gln Ser Ser Thr Val Ala Ile Asp
```

```
    175                 180                 185

cac tca ctg tca tta gca ggt gag agg act tgg gct gaa act atg ggt      627
His Ser Leu Ser Leu Ala Gly Glu Arg Thr Trp Ala Glu Thr Met Gly
190                 195                 200                 205

ctg aat acc gcc gat acg gcc cgg ctc aac gca aat att cgg tac gtc      675
Leu Asn Thr Ala Asp Thr Ala Arg Leu Asn Ala Asn Ile Arg Tyr Val
                210                 215                 220

aac aca ggg act gct cct ata tat aac gtg ctg cct acg aca agt ctt      723
Asn Thr Gly Thr Ala Pro Ile Tyr Asn Val Leu Pro Thr Thr Ser Leu
            225                 230                 235

gtc ctg ggc aaa cag cag acc ctc gca acc att aag gca aag gaa aat      771
Val Leu Gly Lys Gln Gln Thr Leu Ala Thr Ile Lys Ala Lys Glu Asn
        240                 245                 250

cag ctg agc cag atc ctc gcc cct aac aac tat tat cca tcc aaa aat      819
Gln Leu Ser Gln Ile Leu Ala Pro Asn Asn Tyr Tyr Pro Ser Lys Asn
    255                 260                 265

tta gcc ccc ata gcc ctg aac gcc cag gac gac ttt tcc tct acc ccc      867
Leu Ala Pro Ile Ala Leu Asn Ala Gln Asp Asp Phe Ser Ser Thr Pro
270                 275                 280                 285

ata act atg aat tac aat cag ttc ctg gag ctg gaa aag acg aag cag      915
Ile Thr Met Asn Tyr Asn Gln Phe Leu Glu Leu Glu Lys Thr Lys Gln
                290                 295                 300

ctg aga cta gac acc gat cag gtg tat gga aac ata gcg aca tat aac      963
Leu Arg Leu Asp Thr Asp Gln Val Tyr Gly Asn Ile Ala Thr Tyr Asn
            305                 310                 315

ttt gag aac ggc cgc gtg cgc gtc gac act ggg tca cag tgg tct gaa     1011
Phe Glu Asn Gly Arg Val Arg Val Asp Thr Gly Ser Gln Trp Ser Glu
        320                 325                 330

gtt ctg ccg caa att caa gag aca acc gcc aga att atc ttt aat ggg     1059
Val Leu Pro Gln Ile Gln Glu Thr Thr Ala Arg Ile Ile Phe Asn Gly
    335                 340                 345

aag gac ttg aac ctt gtc gaa cgt aga att gcc gcc gtg cag ccc agt     1107
Lys Asp Leu Asn Leu Val Glu Arg Arg Ile Ala Ala Val Gln Pro Ser
350                 355                 360                 365

gat cca ctc gag acg act aaa ccg gat atg aca ctg aaa gag gct ctg     1155
Asp Pro Leu Glu Thr Thr Lys Pro Asp Met Thr Leu Lys Glu Ala Leu
                370                 375                 380

aag att gcc ttc gga ttc aac gaa cct aat ggc aat ttg cag tat cag     1203
Lys Ile Ala Phe Gly Phe Asn Glu Pro Asn Gly Asn Leu Gln Tyr Gln
            385                 390                 395

ggg aaa gac atc aca gag ttt gat ttc aat ttc gat cag cag act tcc     1251
Gly Lys Asp Ile Thr Glu Phe Asp Phe Asn Phe Asp Gln Gln Thr Ser
        400                 405                 410

caa aat atc aaa aat cag ttg gca gag ctg cag gcc acc aat atc tac     1299
Gln Asn Ile Lys Asn Gln Leu Ala Glu Leu Gln Ala Thr Asn Ile Tyr
    415                 420                 425
```

```
acg gtt ctc gat aaa atc aaa ctt aac gcc aag atg aac ata ttg att      1347
Thr Val Leu Asp Lys Ile Lys Leu Asn Ala Lys Met Asn Ile Leu Ile
430                 435                 440                 445

cga gac aaa cgc ttc cac tac gac cgc aac aat ata gcc gta ggc gct      1395
Arg Asp Lys Arg Phe His Tyr Asp Arg Asn Asn Ile Ala Val Gly Ala
                450                 455                 460

gat gag tct gtc gtc aag gag gct cat agg gaa gtt atc cag agc agt      1443
Asp Glu Ser Val Val Lys Glu Ala His Arg Glu Val Ile Gln Ser Ser
            465                 470                 475

act gaa ggg ctg tta ctt aat atc gac aag gac att cgg aag atc ctg      1491
Thr Glu Gly Leu Leu Leu Asn Ile Asp Lys Asp Ile Arg Lys Ile Leu
        480                 485                 490

tcc ggg tat atc gtg gag atc gag gat acc gag ggc ctg aag gaa gtc      1539
Ser Gly Tyr Ile Val Glu Ile Glu Asp Thr Glu Gly Leu Lys Glu Val
    495                 500                 505

att aac gac cgc tat gat atg ctg cag att tcc agc tta cga cag gac      1587
Ile Asn Asp Arg Tyr Asp Met Leu Gln Ile Ser Ser Leu Arg Gln Asp
510                 515                 520                 525

ggt aag aca ttt att gac ttt aaa aag tat aac gac aag cta ccc ctg      1635
Gly Lys Thr Phe Ile Asp Phe Lys Lys Tyr Asn Asp Lys Leu Pro Leu
                530                 535                 540

tac att tcc aac cca aat tac aaa gtt aat gtg tat gct gta acc aag      1683
Tyr Ile Ser Asn Pro Asn Tyr Lys Val Asn Val Tyr Ala Val Thr Lys
            545                 550                 555

gag aac aca atc atc cag cca agc gag aac ggc gat acc agc aca aat      1731
Glu Asn Thr Ile Ile Gln Pro Ser Glu Asn Gly Asp Thr Ser Thr Asn
        560                 565                 570

gga atc aaa aag atc ctt ata ttt agt aaa aaa ggc tac gag atc ggt      1779
Gly Ile Lys Lys Ile Leu Ile Phe Ser Lys Lys Gly Tyr Glu Ile Gly
    575                 580                 585

tgaggatcc                                                            1788
```

<210> 18

<211> 589

<212> PRT

<213> Artificial Sequence

<220>

<223> Human TPA/synthetic antigen fusion protein

<400> 18

```
Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
1               5                   10                  15
```

```
Ala Val Phe Val Ser Pro Ser Ser Ala Gly Pro Thr Val Pro Asp Arg
            20                  25                  30

Asp Asn Asp Gly Ile Pro Asp Ser Leu Glu Val Glu Gly Tyr Thr Val
        35                  40                  45

Asp Val Lys Asn Lys Arg Thr Phe Leu Ser Pro Trp Ile Ser Asn Ile
    50                  55                  60

His Glu Lys Lys Gly Leu Thr Lys Tyr Lys Ser Ser Pro Glu Lys Trp
65                  70                  75                  80

Ser Thr Ala Ser Asp Pro Tyr Ser Asp Phe Glu Lys Val Thr Gly Arg
                85                  90                  95

Ile Asp Lys Gln Val Ser Pro Glu Ala Arg His Pro Leu Val Ala Ala
            100                 105                 110

Tyr Pro Ile Val His Val Asp Met Glu Asn Ile Ile Leu Ser Lys Asn
        115                 120                 125

Glu Asp Gln Ser Thr Gln Asn Thr Asp Ser Glu Thr Arg Thr Ile Ser
    130                 135                 140

Lys Gln Thr Ser Thr Ser Arg Thr His Thr Ser Glu Val His Gly Asn
145                 150                 155                 160

Ala Glu Val His Ala Ser Phe Phe Asp Ile Gly Gly Ser Val Ser Ala
                165                 170                 175

Gly Phe Ser Asn Ser Gln Ser Ser Thr Val Ala Ile Asp His Ser Leu
            180                 185                 190

Ser Leu Ala Gly Glu Arg Thr Trp Ala Glu Thr Met Gly Leu Asn Thr
        195                 200                 205

Ala Asp Thr Ala Arg Leu Asn Ala Asn Ile Arg Tyr Val Asn Thr Gly
    210                 215                 220

Thr Ala Pro Ile Tyr Asn Val Leu Pro Thr Thr Ser Leu Val Leu Gly
225                 230                 235                 240

Lys Gln Gln Thr Leu Ala Thr Ile Lys Ala Lys Glu Asn Gln Leu Ser
                245                 250                 255

Gln Ile Leu Ala Pro Asn Asn Tyr Tyr Pro Ser Lys Asn Leu Ala Pro
            260                 265                 270
```

```
Ile Ala Leu Asn Ala Gln Asp Asp Phe Ser Ser Thr Pro Ile Thr Met
        275                 280                 285

Asn Tyr Asn Gln Phe Leu Glu Leu Glu Lys Thr Lys Gln Leu Arg Leu
    290                 295                 300

Asp Thr Asp Gln Val Tyr Gly Asn Ile Ala Thr Tyr Asn Phe Glu Asn
305                 310                 315                 320

Gly Arg Val Arg Val Asp Thr Gly Ser Gln Trp Ser Glu Val Leu Pro
                325                 330                 335

Gln Ile Gln Glu Thr Thr Ala Arg Ile Ile Phe Asn Gly Lys Asp Leu
            340                 345                 350

Asn Leu Val Glu Arg Arg Ile Ala Ala Val Gln Pro Ser Asp Pro Leu
        355                 360                 365

Glu Thr Thr Lys Pro Asp Met Thr Leu Lys Glu Ala Leu Lys Ile Ala
    370                 375                 380

Phe Gly Phe Asn Glu Pro Asn Gly Asn Leu Gln Tyr Gln Gly Lys Asp
385                 390                 395                 400

Ile Thr Glu Phe Asp Phe Asn Phe Asp Gln Gln Thr Ser Gln Asn Ile
                405                 410                 415

Lys Asn Gln Leu Ala Glu Leu Gln Ala Thr Asn Ile Tyr Thr Val Leu
            420                 425                 430

Asp Lys Ile Lys Leu Asn Ala Lys Met Asn Ile Leu Ile Arg Asp Lys
        435                 440                 445

Arg Phe His Tyr Asp Arg Asn Asn Ile Ala Val Gly Ala Asp Glu Ser
    450                 455                 460

Val Val Lys Glu Ala His Arg Glu Val Ile Gln Ser Ser Thr Glu Gly
465                 470                 475                 480

Leu Leu Leu Asn Ile Asp Lys Asp Ile Arg Lys Ile Leu Ser Gly Tyr
                485                 490                 495

Ile Val Glu Ile Glu Asp Thr Glu Gly Leu Lys Glu Val Ile Asn Asp
            500                 505                 510

Arg Tyr Asp Met Leu Gln Ile Ser Ser Leu Arg Gln Asp Gly Lys Thr
```

```
        515                 520                 525

Phe Ile Asp Phe Lys Lys Tyr Asn Asp Lys Leu Pro Leu Tyr Ile Ser
    530                 535                 540

Asn Pro Asn Tyr Lys Val Asn Val Tyr Ala Val Thr Lys Glu Asn Thr
545                 550                 555                 560

Ile Ile Gln Pro Ser Glu Asn Gly Asp Thr Ser Thr Asn Gly Ile Lys
                565                 570                 575

Lys Ile Leu Ile Phe Ser Lys Lys Gly Tyr Glu Ile Gly
            580                 585
```

<210> 19

<211> 2418

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic coding region for Human TPA/synthetic antigen fusion protein

<220>

<221> CDS

<222> (13)..(2409)

<223>

<400> 19

```
gatatcgcca cc atg gat gca atg aag aga ggg ctc tgc tgt gtg ctg ctg      51
              Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu
              1               5                   10

ctg tgt gga gca gtc ttc gtt tcg ccc agc gcc ggc ggg cat ggg gac        99
Leu Cys Gly Ala Val Phe Val Ser Pro Ser Ala Gly Gly His Gly Asp
    15                  20                  25

gtt ggc atg cat gtg aaa gaa aag gag aaa aac aag gac gaa aac aag       147
Val Gly Met His Val Lys Glu Lys Glu Lys Asn Lys Asp Glu Asn Lys
30                  35                  40                  45

cgt aaa gac gaa gaa cgt cag aaa aca cag gag gaa cac tta aag gag       195
Arg Lys Asp Glu Glu Arg Gln Lys Thr Gln Glu Glu His Leu Lys Glu
                50                  55                  60

atc atg aag cac ata gta aag att gag gta aaa ggc gaa gag gct gta       243
```

```
Ile Met Lys His Ile Val Lys Ile Glu Val Lys Gly Glu Glu Ala Val
            65                  70                  75

aag aag gag gca gca gaa aaa ctg ttg gag aag gtg cct tct gac gtc      291
Lys Lys Glu Ala Ala Glu Lys Leu Leu Glu Lys Val Pro Ser Asp Val
        80                  85                  90

tta gag atg tat aag gcc atc ggc ggt aag atc tat atc gtg gac gga      339
Leu Glu Met Tyr Lys Ala Ile Gly Gly Lys Ile Tyr Ile Val Asp Gly
    95                  100                 105

gac atc act aaa cac ata tct ctc gaa gct ctc tcc gag gac aag aaa      387
Asp Ile Thr Lys His Ile Ser Leu Glu Ala Leu Ser Glu Asp Lys Lys
110                 115                 120                 125

aag att aaa gac atc tac ggg aag gat gcc tta ttg cac gag cac tac      435
Lys Ile Lys Asp Ile Tyr Gly Lys Asp Ala Leu Leu His Glu His Tyr
                130                 135                 140

gtt tac gca aag gag ggc tat gag ccc gtg ctc gtt att cag agt agt      483
Val Tyr Ala Lys Glu Gly Tyr Glu Pro Val Leu Val Ile Gln Ser Ser
            145                 150                 155

gag gac tac gtc gag aat acc gag aaa gct ctg aat gtg tat tac gag      531
Glu Asp Tyr Val Glu Asn Thr Glu Lys Ala Leu Asn Val Tyr Tyr Glu
        160                 165                 170

atc gga aag att ctg tcc cgg gac atc ctg tcc aaa atc aac cag cca      579
Ile Gly Lys Ile Leu Ser Arg Asp Ile Leu Ser Lys Ile Asn Gln Pro
    175                 180                 185

tac cag aaa ttc ctt gat gtt ctt aac aca atc aaa cag gcg tca gat      627
Tyr Gln Lys Phe Leu Asp Val Leu Asn Thr Ile Lys Gln Ala Ser Asp
190                 195                 200                 205

agc gac ggg cag gat ctt ctg ttt aca aat caa ctc aag gaa cac ccc      675
Ser Asp Gly Gln Asp Leu Leu Phe Thr Asn Gln Leu Lys Glu His Pro
                210                 215                 220

act gat ttc agc gtg gag ttc ctc gag cag aat tct aac gaa gtc cag      723
Thr Asp Phe Ser Val Glu Phe Leu Glu Gln Asn Ser Asn Glu Val Gln
            225                 230                 235

gag gtg ttc gcc aag gca ttt gcg tac tat atc gaa ccc cag cat cgc      771
Glu Val Phe Ala Lys Ala Phe Ala Tyr Tyr Ile Glu Pro Gln His Arg
        240                 245                 250

gat gtg ctc cag ctg tac gcc ccg gag gca ttt aac tac atg gac aaa      819
Asp Val Leu Gln Leu Tyr Ala Pro Glu Ala Phe Asn Tyr Met Asp Lys
    255                 260                 265

ttc aat gaa cag gag att cag ctg tct ctg gag gaa ctg aaa gac cag      867
Phe Asn Glu Gln Glu Ile Gln Leu Ser Leu Glu Glu Leu Lys Asp Gln
270                 275                 280                 285

agg atg ctc tcc cgg tat gaa aag tgg gaa aag atc aaa cag cat tac      915
Arg Met Leu Ser Arg Tyr Glu Lys Trp Glu Lys Ile Lys Gln His Tyr
                290                 295                 300

cag cat tgg tcc gac tcc ctg tca gaa gag ggg cgc ggc ctg ttg aaa      963
Gln His Trp Ser Asp Ser Leu Ser Glu Glu Gly Arg Gly Leu Leu Lys
            305                 310                 315
```

```
aag ttg cag att ccc atc gag cct aag aaa gat gat ata ata cac tct      1011
Lys Leu Gln Ile Pro Ile Glu Pro Lys Lys Asp Asp Ile Ile His Ser
        320                     325                     330

cta agc cag gag gag aag gaa ctc ctg aag cgg ata caa atc gac tca      1059
Leu Ser Gln Glu Glu Lys Glu Leu Leu Lys Arg Ile Gln Ile Asp Ser
    335                     340                     345

tcc gat ttc ctt agc aca gaa gag aag gag ttt cta aaa aaa ctt cag      1107
Ser Asp Phe Leu Ser Thr Glu Glu Lys Glu Phe Leu Lys Lys Leu Gln
350                     355                     360                 365

ata gat att aga gat tca ctg agc gag gaa gag aag gag ctg ctc aac      1155
Ile Asp Ile Arg Asp Ser Leu Ser Glu Glu Glu Lys Glu Leu Leu Asn
                370                     375                     380

cga att caa gtc gat agt tcg aac ccc ttg tca gaa aaa gag aag gaa      1203
Arg Ile Gln Val Asp Ser Ser Asn Pro Leu Ser Glu Lys Glu Lys Glu
            385                     390                     395

ttc ctg aaa aag ttg aag ctc gac atc cag ccg tac gat att aat cag      1251
Phe Leu Lys Lys Leu Lys Leu Asp Ile Gln Pro Tyr Asp Ile Asn Gln
        400                     405                     410

cgg cta caa gac acc ggc ggt ctg att gat agc ccc agc atc aac ctt      1299
Arg Leu Gln Asp Thr Gly Gly Leu Ile Asp Ser Pro Ser Ile Asn Leu
    415                     420                     425

gac gta cgg aag caa tat aag cgc gac att caa aat atc gac gcc cta      1347
Asp Val Arg Lys Gln Tyr Lys Arg Asp Ile Gln Asn Ile Asp Ala Leu
430                     435                     440                 445

tta cat caa tcc ata ggc tcc acg cta tac aat aaa atc tat cta tac      1395
Leu His Gln Ser Ile Gly Ser Thr Leu Tyr Asn Lys Ile Tyr Leu Tyr
                450                     455                     460

gaa aac atg aat att aac cag ctc acc gct aca ctg gga gcg gac ctg      1443
Glu Asn Met Asn Ile Asn Gln Leu Thr Ala Thr Leu Gly Ala Asp Leu
            465                     470                     475

gtc gat agt aca gac aac aca aag ata aac aga ggt att ttc aac gaa      1491
Val Asp Ser Thr Asp Asn Thr Lys Ile Asn Arg Gly Ile Phe Asn Glu
        480                     485                     490

ttc aaa aag aac ttt aag tat tcg atc agc agt aac tat atg att gtt      1539
Phe Lys Lys Asn Phe Lys Tyr Ser Ile Ser Ser Asn Tyr Met Ile Val
    495                     500                     505

gac atc aat gaa cgg ccc gca tta gac aat gag agg ttg aag tgg aga      1587
Asp Ile Asn Glu Arg Pro Ala Leu Asp Asn Glu Arg Leu Lys Trp Arg
510                     515                     520                 525

att caa ctg agt cct gat act agg gcc ggc tat ctg gag aac ggg aaa      1635
Ile Gln Leu Ser Pro Asp Thr Arg Ala Gly Tyr Leu Glu Asn Gly Lys
                530                     535                     540

ctg atc tta cag cga aac atc ggg ctg gag atc aag gat gtg cag att      1683
Leu Ile Leu Gln Arg Asn Ile Gly Leu Glu Ile Lys Asp Val Gln Ile
            545                     550                     555

atc aag cag agc gaa aaa gaa tac att cgc atc gac gcc aag gtg gtg      1731
Ile Lys Gln Ser Glu Lys Glu Tyr Ile Arg Ile Asp Ala Lys Val Val
```

560 565 570

cct aag tca aag atc gat acc aag atc cag gaa gct cag ctc aac att 1779
Pro Lys Ser Lys Ile Asp Thr Lys Ile Gln Glu Ala Gln Leu Asn Ile
575 580 585

aac cag gag tgg aat aaa gct ctt ggt ctg cca aaa tac acc aaa ctt 1827
Asn Gln Glu Trp Asn Lys Ala Leu Gly Leu Pro Lys Tyr Thr Lys Leu
590 595 600 605

atc acc ttt aat gtg cac aac agg tat gcc tct aat atc gtc gag tca 1875
Ile Thr Phe Asn Val His Asn Arg Tyr Ala Ser Asn Ile Val Glu Ser
610 615 620

gca tac ctg att ctc aat gaa tgg aag aac aat att cag tct gac ctg 1923
Ala Tyr Leu Ile Leu Asn Glu Trp Lys Asn Asn Ile Gln Ser Asp Leu
625 630 635

atc aag aag gtc acg aat tat ctg gtg gac gga aat ggc aga ttc gtg 1971
Ile Lys Lys Val Thr Asn Tyr Leu Val Asp Gly Asn Gly Arg Phe Val
640 645 650

ttc acc gac ata act ttg cca aac att gcc gag caa tac act cat cag 2019
Phe Thr Asp Ile Thr Leu Pro Asn Ile Ala Glu Gln Tyr Thr His Gln
655 660 665

gat gaa att tac gag caa gtc cac tcc aaa ggt ctg tat gtt cca gag 2067
Asp Glu Ile Tyr Glu Gln Val His Ser Lys Gly Leu Tyr Val Pro Glu
670 675 680 685

tca aga tcg att ctg ctc cat ggt cca tcc aaa ggg gtt gag ctt cga 2115
Ser Arg Ser Ile Leu Leu His Gly Pro Ser Lys Gly Val Glu Leu Arg
690 695 700

cag gat tct gag gga ttt atc gct gac ttt gga gcc gct gtg gat gac 2163
Gln Asp Ser Glu Gly Phe Ile Ala Asp Phe Gly Ala Ala Val Asp Asp
705 710 715

tac gcc gga tac ctg ttg gat aag cag cag tct gat ctc gtg aca aat 2211
Tyr Ala Gly Tyr Leu Leu Asp Lys Gln Gln Ser Asp Leu Val Thr Asn
720 725 730

agc aaa aaa ttc ata gat att ttc aag gag gaa ggg agt cag ctg act 2259
Ser Lys Lys Phe Ile Asp Ile Phe Lys Glu Glu Gly Ser Gln Leu Thr
735 740 745

tcc tat ggc cgc acg aac gag gct gaa ttt ttt gcg gaa gcc ttt aga 2307
Ser Tyr Gly Arg Thr Asn Glu Ala Glu Phe Phe Ala Glu Ala Phe Arg
750 755 760 765

ctt atg cac agc acc gac cat gct gaa agg ttg aag gtg caa aag aat 2355
Leu Met His Ser Thr Asp His Ala Glu Arg Leu Lys Val Gln Lys Asn
770 775 780

gcc cct aaa acc ttc cag ttc ata aat gac cag atc aag ttc atc atc 2403
Ala Pro Lys Thr Phe Gln Phe Ile Asn Asp Gln Ile Lys Phe Ile Ile
785 790 795

aac tct tgaggatcc 2418
Asn Ser

<210> 20

<211> 799

<212> PRT

<213> Artificial Sequence

<220>

<223> Human TPA/synthetic antigen fusion protein

<400> 20

```
Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
1               5                   10                  15

Ala Val Phe Val Ser Pro Ser Ala Gly Gly His Gly Asp Val Gly Met
            20                  25                  30

His Val Lys Glu Lys Glu Lys Asn Lys Asp Glu Asn Lys Arg Lys Asp
        35                  40                  45

Glu Glu Arg Gln Lys Thr Gln Glu Glu His Leu Lys Glu Ile Met Lys
    50                  55                  60

His Ile Val Lys Ile Glu Val Lys Gly Glu Glu Ala Val Lys Lys Glu
65                  70                  75                  80

Ala Ala Glu Lys Leu Leu Glu Lys Val Pro Ser Asp Val Leu Glu Met
                85                  90                  95

Tyr Lys Ala Ile Gly Gly Lys Ile Tyr Ile Val Asp Gly Asp Ile Thr
            100                 105                 110

Lys His Ile Ser Leu Glu Ala Leu Ser Glu Asp Lys Lys Lys Ile Lys
        115                 120                 125

Asp Ile Tyr Gly Lys Asp Ala Leu Leu His Glu His Tyr Val Tyr Ala
    130                 135                 140

Lys Glu Gly Tyr Glu Pro Val Leu Val Ile Gln Ser Ser Glu Asp Tyr
145                 150                 155                 160

Val Glu Asn Thr Glu Lys Ala Leu Asn Val Tyr Tyr Glu Ile Gly Lys
                165                 170                 175

Ile Leu Ser Arg Asp Ile Leu Ser Lys Ile Asn Gln Pro Tyr Gln Lys
            180                 185                 190
```

Phe Leu Asp Val Leu Asn Thr Ile Lys Gln Ala Ser Asp Ser Asp Gly
195 200 205

Gln Asp Leu Leu Phe Thr Asn Gln Leu Lys Glu His Pro Thr Asp Phe
210 215 220

Ser Val Glu Phe Leu Glu Gln Asn Ser Asn Glu Val Gln Glu Val Phe
225 230 235 240

Ala Lys Ala Phe Ala Tyr Tyr Ile Glu Pro Gln His Arg Asp Val Leu
245 250 255

Gln Leu Tyr Ala Pro Glu Ala Phe Asn Tyr Met Asp Lys Phe Asn Glu
260 265 270

Gln Glu Ile Gln Leu Ser Leu Glu Glu Leu Lys Asp Gln Arg Met Leu
275 280 285

Ser Arg Tyr Glu Lys Trp Glu Lys Ile Lys Gln His Tyr Gln His Trp
290 295 300

Ser Asp Ser Leu Ser Glu Glu Gly Arg Gly Leu Leu Lys Lys Leu Gln
305 310 315 320

Ile Pro Ile Glu Pro Lys Lys Asp Asp Ile Ile His Ser Leu Ser Gln
325 330 335

Glu Glu Lys Glu Leu Leu Lys Arg Ile Gln Ile Asp Ser Ser Asp Phe
340 345 350

Leu Ser Thr Glu Glu Lys Glu Phe Leu Lys Lys Leu Gln Ile Asp Ile
355 360 365

Arg Asp Ser Leu Ser Glu Glu Glu Lys Glu Leu Leu Asn Arg Ile Gln
370 375 380

Val Asp Ser Ser Asn Pro Leu Ser Glu Lys Glu Lys Glu Phe Leu Lys
385 390 395 400

Lys Leu Lys Leu Asp Ile Gln Pro Tyr Asp Ile Asn Gln Arg Leu Gln
405 410 415

Asp Thr Gly Gly Leu Ile Asp Ser Pro Ser Ile Asn Leu Asp Val Arg
420 425 430

Lys Gln Tyr Lys Arg Asp Ile Gln Asn Ile Asp Ala Leu Leu His Gln

435 440 445

Ser Ile Gly Ser Thr Leu Tyr Asn Lys Ile Tyr Leu Tyr Glu Asn Met
450 455 460

Asn Ile Asn Gln Leu Thr Ala Thr Leu Gly Ala Asp Leu Val Asp Ser
465 470 475 480

Thr Asp Asn Thr Lys Ile Asn Arg Gly Ile Phe Asn Glu Phe Lys Lys
485 490 495

Asn Phe Lys Tyr Ser Ile Ser Ser Asn Tyr Met Ile Val Asp Ile Asn
500 505 510

Glu Arg Pro Ala Leu Asp Asn Glu Arg Leu Lys Trp Arg Ile Gln Leu
515 520 525

Ser Pro Asp Thr Arg Ala Gly Tyr Leu Glu Asn Gly Lys Leu Ile Leu
530 535 540

Gln Arg Asn Ile Gly Leu Glu Ile Lys Asp Val Gln Ile Ile Lys Gln
545 550 555 560

Ser Glu Lys Glu Tyr Ile Arg Ile Asp Ala Lys Val Val Pro Lys Ser
565 570 575

Lys Ile Asp Thr Lys Ile Gln Glu Ala Gln Leu Asn Ile Asn Gln Glu
580 585 590

Trp Asn Lys Ala Leu Gly Leu Pro Lys Tyr Thr Lys Leu Ile Thr Phe
595 600 605

Asn Val His Asn Arg Tyr Ala Ser Asn Ile Val Glu Ser Ala Tyr Leu
610 615 620

Ile Leu Asn Glu Trp Lys Asn Asn Ile Gln Ser Asp Leu Ile Lys Lys
625 630 635 640

Val Thr Asn Tyr Leu Val Asp Gly Asn Gly Arg Phe Val Phe Thr Asp
645 650 655

Ile Thr Leu Pro Asn Ile Ala Glu Gln Tyr Thr His Gln Asp Glu Ile
660 665 670

Tyr Glu Gln Val His Ser Lys Gly Leu Tyr Val Pro Glu Ser Arg Ser
675 680 685

Ile Leu Leu His Gly Pro Ser Lys Gly Val Glu Leu Arg Gln Asp Ser
    690   695   700

Glu Gly Phe Ile Ala Asp Phe Gly Ala Ala Val Asp Asp Tyr Ala Gly
705   710   715   720

Tyr Leu Leu Asp Lys Gln Gln Ser Asp Leu Val Thr Asn Ser Lys Lys
                725   730   735

Phe Ile Asp Ile Phe Lys Glu Glu Gly Ser Gln Leu Thr Ser Tyr Gly
            740   745   750

Arg Thr Asn Glu Ala Glu Phe Phe Ala Glu Ala Phe Arg Leu Met His
        755   760   765

Ser Thr Asp His Ala Glu Arg Leu Lys Val Gln Lys Asn Ala Pro Lys
    770   775   780

Thr Phe Gln Phe Ile Asn Asp Gln Ile Lys Phe Ile Ile Asn Ser
785   790   795

<210> 21

<211> 2292

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic coding region

<400> 21
atgaagaaga ggaaggtgct gatccccctg atggccctga gcaccatcct ggtgagcagc 60

accggcaacc tggaggtgat ccaggccgag gtgaagcagg agaacaggct gctgaacgag 120

agcgagagca gcagccaggg cctgctgggc tactacttca gcgacctgaa cttccaggcc 180

cccatggtgg tgaccagcag caccaccggc gacctgagca tccccagcag cgagctggag 240

aacatcccca gcgagaacca gtacttccag agcgccatct ggagcggctt catcaaggtg 300

aagaagagcg acgagtacac cttcgccacc agcgccgaca accacgtgac catgtgggtg 360

gacgaccagg aggtgatcaa caaggccagc aacagcaaca agatcaggct ggagaagggc 420

aggctgtacc agatcaagat ccagtaccag agggagaacc ccaccgagaa gggcctggac 480

ttcaagctgt actggaccga cagccagaac aagaaggagg tgatcagcag cgacaacctg 540

cagctgcccg agctgaagca gaagagcagc aacagcagga agaagaggag caccagcgcc 600

ggccccaccg tgcccgacag ggacaacgac ggcatccccg acagcctgga ggtggagggc 660
tacaccgtgg acgtgaagaa caagaggacc ttcctgagcc cctggatcag caacatccac 720
gagaagaagg gcctgaccaa gtacaagagc agccccgaga agtggagcac cgccagcgac 780
ccctacagcg acttcgagaa ggtgaccggc aggatcgaca agaacgtgag ccccgaggcc 840
aggcaccccc tggtggccgc ctaccccatc gtgcacgtgg acatggagaa catcatcctg 900
agcaagaacg aggaccagag cacccagaac accgacagcg agaccaggac catcagcaag 960
aacaccagca ccagcaggac ccacaccagc gaggtgcacg gcaacgccga ggtgcacgcc 1020
agcttcttcg acatcggcgg cagcgtgagc gccggcttca gcaacagcaa cagcagcacc 1080
gtggccatcg accacagcct gagcctggcc ggcgagagga cctgggccga gaccatgggc 1140
ctgaacaccg ccgacaccgc caggctgaac gccaacatca ggtacgtgaa caccggcacc 1200
gcccccatct acaacgtgct gcccaccacc agcctggtgc tgggcaagaa ccagaccctg 1260
gccaccatca aggccaagga gaaccagctg agccagatcc tggcccccaa caactactac 1320
cccagcaaga acctggcccc catcgccctg aacgcccagg acgacttcag cagcaccccc 1380
atcaccatga actacaacca gttcctggag ctggagaaga ccaagcagct gaggctggac 1440
accgaccagg tgtacggcaa catcgccacc tacaacttcg agaacggcag ggtgagggtg 1500
gacaccggca gcaactggag cgaggtgctg ccccagatcc aggagaccac cgccaggatc 1560
atcttcaacg gcaaggacct gaacctggtg gagaggagga tcgccgccgt gaaccccagc 1620
gacccccctgg agaccaccaa gcccgacatg accctgaagg aggccctgaa gatcgccttc 1680
ggcttcaacg agcccaacgg caacctgcag taccagggca aggacatcac cgagttcgac 1740
ttcaacttcg accagcagac cagccagaac atcaagaacc agctggccga gctgaacgcc 1800
accaacatct acaccgtgct ggacaagatc aagctgaacg ccaagatgaa catcctgatc 1860
agggacaaga ggttccacta cgacaggaac aacatcgccg tgggcgccga cgagagcgtg 1920
gtgaaggagg cccacaggga ggtgatcaac agcagcaccg agggcctgct gctgaacatc 1980
gacaaggaca tcaggaagat cctgagcggc tacatcgtgg agatcgagga caccgagggc 2040
ctgaaggagg tgatcaacga caggtacgac atgctgaaca tcagcagcct gaggcaggac 2100
ggcaagacct tcatcgactt caagaagtac aacgacaagc tgcccctgta catcagcaac 2160
cccaactaca aggtgaacgt gtacgccgtg accaaggaga acaccatcat caaccccagc 2220
gagaacggcg acaccagcac caacggcatc aagaagatcc tgatcttcag caagaagggc 2280
tacgagatcg gc 2292

<210> 22

<211> 2427

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic coding region

<400> 22

```
atgaacatca agaaggagtt catcaaggtg atcagcatga gctgcctggt gaccgccatc      60
accctgagcg gccccgtgtt catccccctg gtgcagggcg ccggcggcca cggcgacgtg     120
ggcatgcacg tgaaggagaa ggagaagaac aaggacgaga acaagaggaa ggacgaggag     180
aggaacaaga cccaggagga gcacctgaag gagatcatga agcacatcgt gaagatcgag     240
gtgaagggcg aggaggccgt gaagaaggag gccgccgaga agctgctgga gaaggtgccc     300
agcgacgtgc tggagatgta caaggccatc ggcggcaaga tctacatcgt ggacggcgac     360
atcaccaagc acatcagcct ggaggccctg agcgaggaca agaagaagat caaggacatc     420
tacggcaagg acgccctgct gcacgagcac tacgtgtacg ccaaggaggg ctacgagccc     480
gtgctggtga tccagagcag cgaggactac gtggagaaca ccgagaaggc cctgaacgtg     540
tactacgaga tcggcaagat cctgagcagg gacatcctga gcaagatcaa ccagccctac     600
cagaagttcc tggacgtgct gaacaccatc aagaacgcca gcgacagcga cggccaggac     660
ctgctgttca ccaaccagct gaaggagcac cccaccgact tcagcgtgga gttcctggag     720
cagaacagca acgaggtgca ggaggtgttc gccaaggcct tcgcctacta catcgagccc     780
cagcacaggg acgtgctgca gctgtacgcc cccgaggcct tcaactacat ggacaagttc     840
aacgagcagg agatcaacct gagcctggag gagctgaagg accagaggat gctgagcagg     900
tacgagaagt gggagaagat caagcagcac taccagcact ggagcgacag cctgagcgag     960
gagggcaggg gcctgctgaa gaagctgcag atccccatcg agcccaagaa ggacgacatc    1020
atccacagcc tgagccagga ggagaaggag ctgctgaaga ggatccagat cgacagcagc    1080
gacttcctga gcaccgagga gaaggagttc ctgaagaagc tgcagatcga catcagggac    1140
agcctgagcg aggaggagaa ggagctgctg aacaggatcc aggtggacag cagcaacccc    1200
ctgagcgaga aggagaagga gttcctgaag aagctgaagc tggacatcca gccctacgac    1260
atcaaccaga ggctgcagga caccggcggc ctgatcgaca gccccagcat caacctggac    1320
gtgaggaagc agtacaagag ggacatccag aacatcgacg ccctgctgca ccagagcatc    1380
ggcagcaccc tgtacaacaa gatctacctg tacgagaaca tgaacatcaa caacctgacc    1440
gccaccctgg gcgccgacct ggtggacagc accgacaaca ccaagatcaa caggggcatc    1500
ttcaacgagt tcaagaagaa cttcaagtac agcatcagca gcaactacat gatcgtggac    1560
```

atcaacgaga ggcccgccct ggacaacgag aggctgaagt ggaggatcca gctgagcccc 1620

gacaccaggg ccggctacct ggagaacggc aagctgatcc tgcagaggaa catcggcctg 1680

gagatcaagg acgtgcagat catcaagcag agcgagaagg agtacatcag gatcgacgcc 1740

aaggtggtgc ccaagagcaa gatcgacacc aagatccagg aggcccagct gaacatcaac 1800

caggagtgga acaaggccct gggcctgccc aagtacacca agctgatcac cttcaacgtg 1860

cacaacaggt acgccagcaa catcgtggag agcgcctacc tgatcctgaa cgagtggaag 1920

aacaacatcc agagcgacct gatcaagaag gtgaccaact acctggtgga cggcaacggc 1980

aggttcgtgt tcaccgacat caccctgccc aacatcgccg agcagtacac ccaccaggac 2040

gagatctacg agcaggtgca cagcaagggc ctgtacgtgc ccgagagcag gagcatcctg 2100

ctgcacggcc ccagcaaggg cgtggagctg aggaacgaca gcgagggctt catccacgag 2160

ttcggccacg ccgtggacga ctacgccggc tacctgctgg acaagaacca gagcgacctg 2220

gtgaccaaca gcaagaagtt catcgacatc ttcaaggagg agggcagcaa cctgaccagc 2280

tacggcagga ccaacgaggc cgagttcttc gccgaggcct tcaggctgat gcacagcacc 2340

gaccacgccg agaggctgaa ggtgcagaag aacgccccca agaccttcca gttcatcaac 2400

gaccagatca agttcatcat caacagc 2427

<210> 23

<211> 2295

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic coding region

<400> 23

atgaaaaagc gtaaggtgct gatccctctc atggcgctgt caactatctt ggtgagtagc 60

accggcaacc tggaagtgat ccaagccgaa gtgaagcaag aaaatcgact tctgaacgag 120

agcgaaagtt catcacaggg tcttctcgga tactacttca gtgacttgaa tttccaagca 180

ccaatggtgg tgactagtag caccaccggc gatttgagca ttcccagctc tgagttggag 240

aacattccca gcgaaaatca gtacttccag tctgctatct ggtccggatt cattaaggtt 300

aaaaagtccg acgaatatac atttgctacc tcggcggata accatgtgac aatgtgggtg 360

gacgaccagg aagtgatcaa caaggcttca aactctaata aaatccggct cgagaagggg 420

aggctctacc agatcaaaat tcagtaccag cgggaaaacc ctacagaaaa aggactcgat 480

ttcaagctgt actggacaga tagccaaaac aagaaagaag ttatcagctc agacaatctg 540
cagttacccg agctcaagca gaagagttct aattctagga agaaaagatc tacatccgca 600
gggccaactg tgcccgacag agacaatgat ggaatccctg atagtctaga ggttgaggga 660
tacacggtag atgtcaagaa caaaaggact tttctctcgc cttggatatc aaatatccat 720
gagaagaagg ggcttaccaa gtacaagtcc tcccccgaga agtggtctac cgcttccgat 780
ccatatagcg atttcgagaa ggtcacaggc cggatcgata aaaatgtgtc tccagaggct 840
agacaccccc tggtagcagc ctacccgatt gtacacgtgg acatggagaa catcattcta 900
agcaaaaacg aggaccagtc cacacaaaac actgactccg agacccgcac catatctaaa 960
aacaccagta cttcaaggac ccacacctct gaagtgcacg gcaatgcgga agtccatgca 1020
tcgtttttcg atattggtgg atccgtgtca gccggcttta gcaatagcaa ctcctcgacg 1080
gttgccattg accactcact gtcattagca ggtgagagga cttgggctga aactatgggt 1140
ctgaataccg ccgatacggc ccggctcaac gcaaatattc ggtacgtcaa cacagggact 1200
gctcctatat ataacgtgct gcctacgaca agtcttgtcc tgggcaaaaa tcagaccctc 1260
gcaaccatta aggcaaagga aaatcagctg agccagatcc tcgcccctaa caactattat 1320
ccatccaaaa atttagcccc catagccctg aacgcccagg acgacttttc ctctaccccc 1380
ataactatga attacaatca gttcctggag ctggaaaaga cgaagcagct gagactagac 1440
accgatcagg tgtatggaaa catagcgaca tataactttg agaacggccg cgtgcgcgtc 1500
gacactgggt caaactggtc tgaagttctg ccgcaaattc aagagacaac cgccagaatt 1560
atctttaatg ggaaggactt gaaccttgtc gaacgtagaa ttgccgccgt gaaccccagt 1620
gatccactcg agacgactaa accggatatg acactgaaag aggctctgaa gattgccttc 1680
ggattcaacg aacctaatgg caatttgcag tatcagggga aagacatcac agagtttgat 1740
ttcaatttcg atcagcagac ttcccaaaat atcaaaaatc agttggcaga gctgaatgcc 1800
accaatatct acacggttct cgataaaatc aaacttaacg ccaagatgaa catattgatt 1860
cgagacaaac gcttccacta cgaccgcaac aatatagccg taggcgctga tgagtctgtc 1920
gtcaaggagg ctcataggga agttatcaac agcagtactg aagggctgtt acttaatatc 1980
gacaaggaca ttcggaagat cctgtccggg tatatcgtgg agatcgagga taccgagggc 2040
ctgaaggaag tcattaacga ccgctatgat atgctgaaca tttccagctt acgacaggac 2100
ggtaagacat ttattgactt taaaaagtat aacgacaagc taccoctgta catttccaac 2160
ccaaattaca aagttaatgt gtatgctgta accaaggaga acacaatcat caatccaagc 2220
gagaacggcg ataccagcac aaatggaatc aaaaagatcc ttatatttag taaaaaaggc 2280
tacgagatcg gttga 2295

<210> 24

<211> 2292

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic coding region

<400> 24

```
atgaaaaaga ggaaggtgct gatccctctc atggccctgt ctaccatcct ggttagtagc      60
acagggaacc tggaagtgat tcaggccgag gttaagcaag agaataggct gctcaacgag     120
tcagaatctt cgtcacaggg attattgggt tactattttt cggacctgaa tttccaggcc     180
ccaatggtcg ttacaagctc cacaacaggc gacctgtcta tccccagctc cgaattggag     240
aacatcccta gcgagaacca atactttcaa agcgctattt ggtcaggctt cataaaagtg     300
aagaagtctg acgaatacac gtttgcaaca tctgccgata accacgtcac tatgtgggtc     360
gatgaccagg aagtcatcaa caaggctagt aatagcaaca aaatcagact ggagaaaggg     420
agattgtacc agatcaagat ccagtaccaa cgggaaaacc caacagagaa gggcctcgat     480
tttaaactgt attggactga ctctcagaat aagaaggaag tgattagcag cgacaattta     540
caattacccg agttgaaaca gaagagctct aattcaagga aaaagagatc tacctccgcc     600
ggaccaacag ttccagatag ggataatgat ggaatccctg actcactgga ggtcgagggt     660
tacaccgtgg acgtgaaaaa caaacgcact ttcctatcac cctggatctc caacattcac     720
gagaagaagg gtctgactaa gtacaaatcc agcccagaga aatggagcac cgcaagtgat     780
ccttatagtg acttcgagaa ggtgacgggc cggatagaca agaacgtatc acccgaagct     840
cgtcatcctc tggtcgccgc ctaccctatt gtgcatgtgg acatggaaaa catcatcctg     900
agtaagaacg aagaccagag cactcagaac accgactccg agacacgaac gatatctaag     960
aatacatcca cctcacgcac tcataccagc gaagtgcacg gtaacgctga agtgcacgcg    1020
tccttcttcg acatcggcgg gtccgtgtcc gctggatttt ccaactccaa ctcttcgacc    1080
gtagctattg accacagcct gagccttgcc ggagaaagga catgggcgga gactatgggc    1140
ctgaatacgg ctgatacggc acggctcaat gccaacatca gatacgtgaa caccgggaca    1200
gcccctattt acaatgtgct cccaaccaca tcactcgtac tgggaaaaaa ccagacccta    1260
gctactatta aagcgaaaga aaatcagttg tcacagatac tggcacccaa caattattat    1320
ccaagcaaaa acctggcacc catcgcactc aatgcgcagg atgactttag tagtacaccc    1380
attacaatga actacaatca gttccttgag ctcgagaaga ccaagcaact gagactcgac    1440
```

actgaccagg tgtatggtaa tatcgccacc tacaacttcg aaaacggaag ggtgcgggta 1500

gatacgggct ccaactggtc ggaggtgtta cctcagatcc aggaaaccac cgcccgcatt 1560

attttcaacg gcaaggatct aaatctcgtt gagcgccgga tagcagctgt caatcccagt 1620

gatccgctcg aaaccaccaa gcccgacatg actttgaagg aagctcttaa gattgccttt 1680

gggtttaacg agcccaatgg gaatctgcag taccagggca aagatattac cgagtttgat 1740

tttaactttg atcagcagac aagccagaac attaagaatc aattagccga gctgaatgcc 1800

actaacatct atactgtttt ggacaagatt aaacttaatg caaagatgaa tatactaata 1860

cgagataagc gcttccatta tgatcgaaat aacatcgcag ttggcgccga cgaatcagtt 1920

gtgaaggagg cccacagaga ggtcattaat tcctctacag agggtcttct tctgaatatc 1980

gacaaggaca tacgtaagat cctgagcgga tatattgtag aaatcgaaga tactgagggg 2040

ctgaaagagg tcatcaacga ccgctatgat atgcttaaca tctctagttt gcggcaagat 2100

ggaaagactt tcattgattt caagaaatac aacgataaac tcccgctgta tatctccaac 2160

ccaaattata aggtgaatgt gtacgctgtc accaaagaga ataccattat taacccgtct 2220

gagaatggcg acacctccac gaatgggata aaaaaaatcc ttatcttcag taaaaaaggc 2280

tacgagatcg gg 2292

<210> 25

<211> 2292

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic coding region

<400> 25
atgaaaaaaa ggaaggtgtt gatccccctt atggcccttt ccaccatctt ggtatcctca 60

accggcaacc tggaggttat tcaagccgaa gtgaagcagg aaaatagact gctgaacgaa 120

tccgaatcta gttctcaggg tctgctgggc tactatttta gcgacctcaa tttccaggca 180

ccaatggtcg tgacttcgag caccacaggc gacttgagca ttccctcttc cgaactcgag 240

aacataccaa gcgagaatca gtattttcag tccgcaatct ggtcgggttt tatcaaagta 300

aaaaagagcg acgaatacac tttcgctacg tcagccgata atcatgtgac catgtgggtg 360

gatgaccaag aggtcatcaa taaggcgagt aactctaaca agattcgact ggaaaaggga 420

cgcctctatc agattaagat tcagtaccag cgtgagaacc ccactgaaaa gggtctggac 480

tttaagctgt attggacgga tagtcagaat aaaaaggagg tgatcagttc agacaacttg 540

caattgcctg agctgaaaca gaagtccagc aactctcgga agaagcgcag tactagcgct 600

ggcccaacag tccccgaccg cgacaatgat gggattcccg attctttgga agtggaggga 660

tacacagtgg acgtgaagaa caagagaaca ttcctgagtc catggattag taatatccat 720

gagaaaaaag gtctaaccaa atacaaaagc agcccagaga agtggtcaac agcatcggat 780

ccttactccg atttcgagaa agttactggc aggattgaca agaacgtatc tccggaggcc 840

aggcatcctc tcgtcgccgc ttacccgatc gtccacgtcg acatggagaa catcatcctg 900

agtaagaatg aggatcaaag cactcagaat actgattccg agacacggac aatcagtaag 960

aatacctcaa cgagcaggac acacacctct gaagttcacg gcaatgccga ggtgcacgct 1020

tcattcttcg atatcggagg atccgtgagc gcgggcttca gcaactctaa ctcttccact 1080

gtagcgatcg atcatagcct ctccctagcc ggagagcgga catgggctga gaccatgggg 1140

ttgaatactg cagatacagc aagactgaac gccaatatta ggtatgtgaa tacaggtacc 1200

gcccccatct acaacgtcct tcctaccacc tcactggtgt taggcaaaaa tcagaccctc 1260

gccaccatta aggcaaaaga gaatcaactc tcacagatac tggccccaaa caactattac 1320

ccatctaaga atttagctcc cattgcttta aacgcccagg acgattttag ctcaacgcct 1380

atcaccatga attataacca gttcctggaa ctggaaaaaa ctaagcagct ccgcctggac 1440

accgatcagg tgtatggcaa catcgccaca tacaatttcg aaaatgggcg cgttcgggtg 1500

gacaccgggt ccaactggag tgaagtccta ccccaaatcc aggaaaccac tgctcgaatc 1560

atcttcaatg gaaaagacct gaatcttgtg gagcggcgaa tcgccgctgt gaatccttcc 1620

gaccctctgg aaacgacgaa gcccgacatg actttgaaag aggcgctaaa aatcgctttt 1680

ggatttaatg aaccgaacgg caacttacaa tatcaaggga aggacattac cgagttcgac 1740

tttaactttg atcagcagac ctcgcagaac ataaagaatc agctcgctga gctgaacgca 1800

acgaatatat acaccgtcct ggacaaaatt aagcttaacg ccaagatgaa catcctcatt 1860

agagacaaga gatttcacta cgataggaat aacattgccg ttggagccga tgagtctgtg 1920

gtgaaagagg cacaccgcga ggtcattaac tccagcactg aagggctgct gctgaacatt 1980

gacaaggata ttagaaaaat cctgagcggg tacatcgttg agatcgaaga taccgaggga 2040

cttaaggaag ttataaacga ccgttatgac atgttaaaca tatcaagcct ccggcaggac 2100

ggtaagacat ttatagattt caagaaatac aacgataagc ttcctcttta catctcaaat 2160

cccaactata aggtgaatgt ttatgcagta acaaaagaaa atacaattat taatccatcc 2220

gagaacggcg atacatctac taacgggata aaaaaaatcc tcatcttctc caagaaaggc 2280

tacgagatag gg 2292

<210> 26

<211> 2430

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic coding region

<400> 26

```
atgaacatca aaaaagagtt tataaaggtg attagcatga gctgcctggt cactgccatt      60
accctgagtg gcccagtgtt tatccctctc gtccagggcg ccggcgggca tggggacgtt     120
ggcatgcatg tgaaagaaaa ggagaaaaac aaggacgaaa acaagcgtaa agacgaagaa     180
cgtaataaaa cacaggagga acacttaaag gagatcatga agcacatagt aaagattgag     240
gtaaaaggcg aagaggctgt aaagaaggag gcagcagaaa aactgttgga gaaggtgcct     300
tctgacgtct tagagatgta taaggccatc ggcggtaaga tctatatcgt ggacggagac     360
atcactaaac acatatctct cgaagctctc tccgaggaca agaaaaagat taaagacatc     420
tacgggaagg atgccttatt gcacgagcac tacgtttacg caaaggaggg ctatgagccc     480
gtgctcgtta ttcagagtag tgaggactac gtcgagaata ccgagaaagc tctgaatgtg     540
tattacgaga tcggaaagat tctgtcccgg gatatcctgt ccaaaatcaa ccagccatac     600
cagaaattcc ttgatgttct taacacaatc aaaaacgcgt cagatagcga cgggcaggat     660
cttctgttta caaatcaact caaggaacac cccactgatt tcagcgtgga gttcctcgag     720
cagaattcta acgaagtcca ggaggtgttc gccaaggcat ttgcgtacta tatcgaaccc     780
cagcatcgcg atgtgctcca gctgtacgcc ccggaggcat ttaactacat ggacaaattc     840
aatgaacagg agattaatct gtctctggag gaactgaaag accagaggat gctctcccgg     900
tatgaaaagt gggaaaagat caaacagcat taccagcatt ggtccgactc cctgtcagaa     960
gaggggcgcg gcctgttgaa aaagttgcag attcccatcg agcctaagaa agatgatata    1020
atacactctc taagccagga ggagaaggaa ctcctgaagc ggatacaaat cgactcatcc    1080
gatttcctta gcacagaaga gaaggagttt ctaaaaaaac ttcagataga tattagagat    1140
tcactgagcg aggaagagaa ggagctgctc aaccgaattc aagtcgatag ttcgaacccc    1200
ttgtcagaaa aagagaagga attcctgaaa aagttgaagc tcgacatcca gccgtacgat    1260
attaatcagc ggctacaaga caccggcggt ctgattgata gccccagcat caaccttgac    1320
gtacggaagc aatataagcg cgacattcaa aatatcgacg ccctattaca tcaatccata    1380
ggatccacgc tatacaataa aatctatcta tacgaaaaca tgaatattaa caatctcacc    1440
```

gctacactgg gagcggacct ggtcgatagt acagacaaca caaagataaa cagaggtatt 1500

ttcaacgaat tcaaaaagaa ctttaagtat tcgatcagca gtaactatat gattgttgac 1560

atcaatgaac ggcccgcatt agacaatgag aggttgaagt ggagaattca actgagtcct 1620

gatactaggg ccggctatct ggagaacggg aaactgatct tacagcgaaa catcgggctg 1680

gagatcaagg atgtgcagat tatcaagcag agcgaaaaag aatacattcg catcgacgcc 1740

aaggtggtgc ctaagtcaaa gatcgatacc aagatccagg aagctcagct caacattaac 1800

caggagtgga ataaagctct tggtctgcca aaatacacca aacttatcac ctttaatgtg 1860

cacaacaggt atgcctctaa tatcgtcgag tcagcatacc tgattctcaa tgaatggaag 1920

aacaatattc agtctgacct gatcaagaag gtcacgaatt atctggtgga cggaaatggc 1980

agattcgtgt tcaccgacat aactttgcca aacattgccg agcaatacac tcatcaggat 2040

gaaatttacg agcaagtcca ctccaaaggt ctgtatgttc cagagtcaag atcgattctg 2100

ctccatggtc catccaaagg ggttgagctt cgaaacgatt ctgagggatt tatccacgag 2160

tttggacacg ctgtggatga ctacgccgga tacctgttgg ataagaatca gtctgatctc 2220

gtgacaaata gcaaaaaatt catagatatt ttcaaggagg aagggagtaa cctgacttcc 2280

tatggccgca cgaacgaggc tgaatttttt gcggaagcct ttagacttat gcacagcacc 2340

gaccatgctg aaaggttgaa ggtgcaaaag aatgccccta aaaccttcca gttcataaat 2400

gaccagatca agttcatcat caactcttga 2430

<210> 27

<211> 2427

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic coding region

<400> 27

atgaatatta aaaaggagtt tattaaggtt atctctatgt cctgcttggt gacagcgata 60

acactgtcag gaccagtgtt catacccctt gtccaggggg ccggcggtca tggcgatgta 120

ggtatgcatg tgaaagagaa ggaaaaaaat aaagacgaga acaagaggaa ggacgaggaa 180

aggaataaga cccaagagga gcacctgaaa gagatcatga agcatattgt gaaaatcgag 240

gtgaaggggg aagaggccgt gaaaaaagaa gcagctgaga agctgctaga gaaagtgcct 300

tctgacgtcc tcgagatgta caaagcaatc ggcggcaaga tttacattgt tgatggtgac 360

ataacaaagc atatttctct ggaggcctta agtgaagata agaaaaaaat caaagacatt 420
tacggaaagg acgccctcct gcacgagcat tatgtctacg ctaaagaagg ctacgaaccc 480
gtgctcgtca tccagtcatc tgaggattat gtggagaaca ccgaaaaagc tttgaacgtc 540
tattacgaaa ttgggaaaat tctgtctaga gacattctca gcaagatcaa ccagccatac 600
caaaaattcc tagacgttct aaatacgatc aagaatgcca gtgactccga cgggcaggat 660
ctgttgttta cgaaccagct taaggagcat cctaccgatt tttctgtcga attccttgag 720
cagaactcca atgaagttca agaggtcttt gctaaggctt tcgcgtacta catcgagcct 780
cagcaccggg acgtgctgca gctctacgcc cctgaagctt tcaattatat ggacaagttc 840
aatgaacagg aaattaacct gagtttagaa gaactgaaag accaaagaat gttgtccaga 900
tacgagaagt gggagaagat caagcagcac tatcagcact ggtccgattc ccttagcgaa 960
gaagggcgcg ggctgcttaa aaagctgcag attccgatcg agccgaagaa agacgatata 1020
attcattcac tgagccagga ggaaaaggag ctcctcaaac ggatccagat cgactcgtcc 1080
gatttcctat ccacagagga aaaggaattt cttaaaaaac tccagattga tatacgggac 1140
tcattatctg aggaggaaaa ggaactcctg aaccggatcc aggtcgatag tagcaacccc 1200
ctgtcagaaa aggaaaagga gtttctcaaa aaacttaagc tggatatcca accatacgac 1260
atcaaccagc gactgcagga tactggaggc ttgatagatt ctccctccat aaacctggac 1320
gtgaggaagc agtataagag ggatatccag aatatcgatg ccctgctgca ccaatctatc 1380
ggaagtactc tttacaacaa aatctatctg tatgagaaca tgaatattaa taacctgact 1440
gctaccttgg gcgccgacct ggtggactcg acggacaaca ccaaaatcaa ccgggggatc 1500
ttcaatgaat ttaagaaaaa tttcaagtac tccatttcca gtaattatat gatagttgat 1560
atcaacgagc ggccagcact ggacaacgag agattaaagt ggcgaattca actgagtccc 1620
gatacacgcg ccggttacct cgagaacggt aagttgatct tgcagcgaaa catcggactc 1680
gagattaagg atgtacagat catcaagcag agcgagaagg agtacattcg tatcgacgct 1740
aaagtggtac caaaaagcaa aattgacacc aagatacagg aggcacagct gaacataaat 1800
caagaatgga ataaagccct cggtctgcct aagtatacta agctaatcac ctttaacgtg 1860
cacaatagat atgccagcaa tattgtcgag agcgcatacc taattctgaa cgaatggaaa 1920
aacaatatcc aaagcgacct gataaaaaag gtgactaatt atctcgttga tggcaatggc 1980
cgcttcgttt tcaccgatat tactctcccc aacatcgcag aacagtacac tcatcaggac 2040
gagatttatg aacaggtgca cagtaagggg ctgtatgtcc ctgagagccg ctctatcctt 2100
cttcacggac cctcaaaggg cgtagagtta aggaatgaca gcgaggggtt cattcacgag 2160
tttggccacg cagtggatga ttacgctgga tatctcctgg ataagaacca gtccgacctg 2220

gtgacaaact caaagaagtt catcgatatc tttaaggagg aaggcagcaa tttgaccagc 2280

tacggacgca caaatgaggc cgagtttttt gccgaagcgt tccgtttgat gcattcaacc 2340

gaccacgcgg aaagactgaa agtgcagaaa aacgccccaa agacattcca gtttattaac 2400

gaccaaatca agttcataat caattcg 2427

<210> 28

<211> 2427

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic coding region

<400> 28

atgaatatta agaaagaatt cattaaagtc attagcatga gctgcctagt caccgccatc 60

accctctctg gcccagtgtt tattccactg gtacagggcg caggcgggca tggcgacgtg 120

ggaatgcatg ttaaggagaa ggaaaaaaat aaagatgaaa acaaacgcaa ggacgaagaa 180

cggaacaaga cccaggagga gcacctgaaa gagatcatga aacacattgt gaaaatcgaa 240

gttaaaggtg aagaggccgt gaagaaggaa gccgcggaga aactgctgga aaaggtcccg 300

tcggacgtac ttgaaatgta caaggcaatt ggtggcaaaa tctacattgt ggacggggac 360

attaccaagc acataagcct ggaagcactc agcgaggaca agaagaaaat aaaggacatt 420

tacggaaagg acgctctgct ccacgagcac tatgtctacg cgaaggaggg gtacgagccc 480

gtgttggtga tacagagttc cgaggactat gttgaaaata ctgaaaaagc cctcaacgtg 540

tactatgaga ttggtaagat cttgtctaga gacattctca gcaagattaa ccagccctac 600

cagaaattcc tggatgtcct gaacacgatt aagaatgcct cagacagcga tggacaggac 660

cttctgttta ccaatcagct taaagagcac cctaccgatt tctccgtgga attccttgag 720

cagaattcaa atgaggtgca agaggtcttc gctaaggcct ttgcctacta tatcgagccc 780

cagcatcgag acgtgctaca gttgtatgca ccagaagcct ttaactacat ggacaagttc 840

aatgagcaag agatcaactt atcactggag gagctgaagg atcaacgcat gctgtctcgg 900

tatgaaaaat gggagaaaat aaagcagcat taccagcatt ggagcgactc cctgtctgaa 960

gagggtcgcg gcctcctgaa aaagctgcag attcctatcg agcctaaaaa agatgatata 1020

attcactcac tgtcccagga agagaaggag ctgcttaagc ggatccagat agattccagt 1080

gacttcttaa gcacggaaga gaaggaattt ctgaaaaaat tgcagatcga tatccgtgat 1140

tccctgtcag aggaggaaaa agaactgctg aaccggattc aggtggattc ctccaatcca 1200

ctgtccgaaa aggagaagga gtttttaaag aaactgaagc tcgatatcca gccttacgat 1260

atcaatcaaa ggttgcagga cacaggggga ttgattgact cgcccagtat caacttggac 1320

gtgagaaagc agtacaaacg agacatacag aatatcgatg ctctcctgca tcaaagcatt 1380

gggtctacac tttacaacaa gatttatttg tacgaaaaca tgaatatcaa taatttgacc 1440

gccactttag gtgctgatct cgtggactcc actgataaca caaagattaa taggggcatt 1500

ttcaatgagt tcaaaaagaa tttcaagtat tctattagct ctaactatat gattgttgat 1560

atcaacgaac gaccagccct agacaacgaa cggctaaagt ggcggatcca gctatcaccg 1620

gacaccagag ctggctacct cgaaaatggg aagctcattc tccagaggaa tatcggactg 1680

gagataaagg acgttcaaat aataaaacaa agcgagaagg agtacataag gatcgatgct 1740

aaggtcgtgc ccaaaagtaa gatcgatact aagattcaag aggctcaact gaatatcaat 1800

caggagtgga ataaggctct cgggctgccc aagtacacaa agctgatcac attcaatgtt 1860

cataacagat acgcgtctaa catcgtcgag tccgcgtatc tgatccttaa tgaatggaag 1920

aacaacatcc agtccgatct catcaagaaa gtcactaact acttagtaga tgggaacgga 1980

cgctttgtgt ttacagatat cacactccca aacatcgctg aacagtatac gcaccaagat 2040

gaaatctatg agcaggtgca cagtaagggc ctgtacgtgc ctgagagtag aagtatcctt 2100

ctgcacggcc ccagcaaagg cgtagagctt cgtaacgatt cggagggatt catacatgag 2160

tttgggcacg cagtcgacga ctatgccggt tatcttctcg acaaaaacca gagcgacttg 2220

gttaccaaca gtaagaaatt tatcgatatc ttcaaagaag agggctctaa cctaacatca 2280

tatggaagga ctaacgaggc agaattcttt gccgaagcct tccgcctgat gcattcaacc 2340

gaccacgcag agagactgaa agtgcagaaa aacgccccta agactttcca gtttattaat 2400

gaccagatca aatttatcat caactct 2427

<210> 29

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 29
gagcttgata tcgccaccat ggatgc 26

<210> 30

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 30
ccaccaatat ccgatgcatg gacttccgc 29

<210> 31

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 31
cttgaaggat cctcaaccga tctcgtag 28

<210> 32

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Syntheric PCR Primer

<400> 32
ccatgcatcg gatattggtg gctccgtgtc 30

<210> 33

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 33
gtggacgacc aggaagtgat c 21

<210> 34

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 34
ggctatctgt ccagtacagc ttgaa 25

<210> 35

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 35
ccgtgctcgt tattcagagt 20

<210> 36

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 36
ccttctcttc tgtgctaagg 20

<210> 37

<211> 34

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 37
gaacctggat ccctacacca ccttggcgtc gatg 34

<210> 38

<211> 35

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 38
gctaatggat cctcaaaatg ccttggcgaa cacct 35

<210> 39

<211> 876

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic coding region for Human/B. anthracis
antigen fusion protein

<220>

<221> CDS

<222> (13)..(870)

<223>

<400> 39

```
gatatcgcca cc atg gat gca atg aag aga ggg ctc tgc tgt gtg ctg ctg      51
              Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu
              1               5                   10

ctg tgt gga gca gtc ttc gtt tcg ccc agc gcc ggc ggg cat ggg gac        99
Leu Cys Gly Ala Val Phe Val Ser Pro Ser Ala Gly Gly His Gly Asp
    15                  20                  25

gtt ggc atg cat gtg aaa gaa aag gag aaa aac aag gac gaa aac aag       147
Val Gly Met His Val Lys Glu Lys Glu Lys Asn Lys Asp Glu Asn Lys
30                  35                  40                  45

cgt aaa gac gaa gaa cgt aat aaa aca cag gag gaa cac tta aag gag       195
Arg Lys Asp Glu Glu Arg Asn Lys Thr Gln Glu Glu His Leu Lys Glu
                50                  55                  60

atc atg aag cac ata gta aag att gag gta aaa ggc gaa gag gct gta       243
Ile Met Lys His Ile Val Lys Ile Glu Val Lys Gly Glu Glu Ala Val
            65                  70                  75

aag aag gag gca gca gaa aaa ctg ttg gag aag gtg cct tct gac gtc       291
Lys Lys Glu Ala Ala Glu Lys Leu Leu Glu Lys Val Pro Ser Asp Val
        80                  85                  90

tta gag atg tat aag gcc atc ggc ggt aag atc tat atc gtg gac gga       339
Leu Glu Met Tyr Lys Ala Ile Gly Gly Lys Ile Tyr Ile Val Asp Gly
    95                  100                 105

gac atc act aaa cac ata tct ctc gaa gct ctc tcc gag gac aag aaa       387
Asp Ile Thr Lys His Ile Ser Leu Glu Ala Leu Ser Glu Asp Lys Lys
110                 115                 120                 125

aag att aaa gac atc tac ggg aag gat gcc tta ttg cac gag cac tac       435
Lys Ile Lys Asp Ile Tyr Gly Lys Asp Ala Leu Leu His Glu His Tyr
                130                 135                 140

gtt tac gca aag gag ggc tat gag ccc gtg ctc gtt att cag agt agt       483
Val Tyr Ala Lys Glu Gly Tyr Glu Pro Val Leu Val Ile Gln Ser Ser
            145                 150                 155

gag gac tac gtc gag aat acc gag aaa gct ctg aat gtg tat tac gag       531
Glu Asp Tyr Val Glu Asn Thr Glu Lys Ala Leu Asn Val Tyr Tyr Glu
        160                 165                 170

atc gga aag att ctg tcc cgg gac atc ctg tcc aaa atc aac cag cca       579
Ile Gly Lys Ile Leu Ser Arg Asp Ile Leu Ser Lys Ile Asn Gln Pro
    175                 180                 185

tac cag aaa ttc ctt gat gtt ctt aac aca atc aaa aac gcg tca gat       627
Tyr Gln Lys Phe Leu Asp Val Leu Asn Thr Ile Lys Asn Ala Ser Asp
190                 195                 200                 205

agc gac ggg cag gat ctt ctg ttt aca aat caa ctc aag gaa cac ccc       675
Ser Asp Gly Gln Asp Leu Leu Phe Thr Asn Gln Leu Lys Glu His Pro
                210                 215                 220

act gat ttc agc gtg gag ttc ctc gag cag aat tct aac gaa gtc cag       723
Thr Asp Phe Ser Val Glu Phe Leu Glu Gln Asn Ser Asn Glu Val Gln
            225                 230                 235

gag gtg ttc gcc aag gca ttt gcg tac tat atc gaa ccc cag cat cgc       771
Glu Val Phe Ala Lys Ala Phe Ala Tyr Tyr Ile Glu Pro Gln His Arg
```

```
        240                 245                 250

gat gtg ctc cag ctg tac gcc ccg gag gca ttt aac tac atg gac aaa      819
Asp Val Leu Gln Leu Tyr Ala Pro Glu Ala Phe Asn Tyr Met Asp Lys
    255                 260                 265

ttc aat gaa cag gag att aat ctg tct ctg gag gaa ctg aaa gac cag      867
Phe Asn Glu Gln Glu Ile Asn Leu Ser Leu Glu Glu Leu Lys Asp Gln
270                 275                 280                 285

tga ggatcc                                                           876
```

<210> 40

<211> 285

<212> PRT

<213> Artificial Sequence

<220>

<223> Human/B. anthracis antigen fusion protein

<400> 40

```
Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
1               5                   10                  15

Ala Val Phe Val Ser Pro Ser Ala Gly Gly His Gly Asp Val Gly Met
            20                  25                  30

His Val Lys Glu Lys Glu Lys Asn Lys Asp Glu Asn Lys Arg Lys Asp
        35                  40                  45

Glu Glu Arg Asn Lys Thr Gln Glu Glu His Leu Lys Glu Ile Met Lys
    50                  55                  60

His Ile Val Lys Ile Glu Val Lys Gly Glu Glu Ala Val Lys Lys Glu
65                  70                  75                  80

Ala Ala Glu Lys Leu Leu Glu Lys Val Pro Ser Asp Val Leu Glu Met
                85                  90                  95

Tyr Lys Ala Ile Gly Gly Lys Ile Tyr Ile Val Asp Gly Asp Ile Thr
            100                 105                 110

Lys His Ile Ser Leu Glu Ala Leu Ser Glu Asp Lys Lys Lys Ile Lys
        115                 120                 125

Asp Ile Tyr Gly Lys Asp Ala Leu Leu His Glu His Tyr Val Tyr Ala
```

130 135 140

Lys Glu Gly Tyr Glu Pro Val Leu Val Ile Gln Ser Ser Glu Asp Tyr
145 150 155 160

Val Glu Asn Thr Glu Lys Ala Leu Asn Val Tyr Tyr Glu Ile Gly Lys
165 170 175

Ile Leu Ser Arg Asp Ile Leu Ser Lys Ile Asn Gln Pro Tyr Gln Lys
180 185 190

Phe Leu Asp Val Leu Asn Thr Ile Lys Asn Ala Ser Asp Ser Asp Gly
195 200 205

Gln Asp Leu Leu Phe Thr Asn Gln Leu Lys Glu His Pro Thr Asp Phe
210 215 220

Ser Val Glu Phe Leu Glu Gln Asn Ser Asn Glu Val Gln Glu Val Phe
225 230 235 240

Ala Lys Ala Phe Ala Tyr Tyr Ile Glu Pro Gln His Arg Asp Val Leu
245 250 255

Gln Leu Tyr Ala Pro Glu Ala Phe Asn Tyr Met Asp Lys Phe Asn Glu
260 265 270

Gln Glu Ile Asn Leu Ser Leu Glu Glu Leu Lys Asp Gln
275 280 285

<210> 41

<211> 35

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 41
ccatacggat cctcactggt ctttcagttc ctcca 35

<210> 42

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 42
ctggagacac ctgtttatcg atcc 24

<210> 43

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 43
ggatcgataa acaggtgtct ccag 24

<210> 44

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 44
gaagtactgg tctgtttaga tatggt 26

<210> 45

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 45
accatatcta aacagaccag tacttc 26

<210> 46

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 46
cgtcgaggac tggctattgc taa 23

<210> 47

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 47
ttagcaatag ccagtcctcg acg 23

<210> 48

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 48
gagggtctgc tgtttgccca gg 22

<210> 49

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 49
cctgggcaaa cagcagaccc tc 22

<210> 50

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 50
cttcagacca ctgtgaccca gtg 23

<210> 51

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 51
cactgggtca cagtggtctg aag 23

<210> 52

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 52
gatcactggg ctgcacggcg g 21

<210> 53

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 53
ccgccgtgca gcccagtgat c 21

<210> 54

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 54
tattggtggc ctgcagctct gc 22

<210> 55

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 55
gcagagctgc aggccaccaa ta 22

<210> 56

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 56
cagtactgct ctggataact tccc 24

<210> 57

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 57
gggaagttat ccagagcagt actg 24

<210> 58

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 58
aagctggaaa tctgcagcat atcat 25

<210> 59

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 59
atgatatgct gcagatttcc agctt 25

<210> 60

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 60
ctcgcttggc tggatgattg tgt 23

<210> 61

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 61
acacaatcat ccagccaagc gag 23

<210> 62

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 62
tcctgtgttt tctgacgttc ttcg 24

<210> 63

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 63
cgaagaacgt cagaaaacac agga 24

<210> 64

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 64
tatctgacgc ctgtttgatt gtgtt 25

<210> 65

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 65
aacacaatca aacaggcgtc agata 25

<210> 66

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 66
ccagagacag ctgaatctcc tgtt 24

<210> 67

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 67
aacaggagat tcagctgtct ctgg 24

<210> 68

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 68
agcggtgagc tggttaatat tcatg 25

<210> 69

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 69
catgaatatt aaccagctca ccgct 25

<210> 70

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 70
cctcagaatc ctgtcgaagc tca 23

<210> 71

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 71
tgagcttcga caggattctg agg 23

<210> 72

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 72
gatcagactg ctgcttatcc aaca 24

<210> 73

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 73
tgttggataa gcagcagtct gatc

<210> 74

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 74
aggaagtcag ctgactccct tcc 23

<210> 75

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 75
ggaagggagt cagctgactt cct 23

<210> 76

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic PCR primer

<400> 76
gcagatctgg atcctcaaga g 21